

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant : Ashkenazi et al.

App. No. : 09/903,925

Filed : July 11, 2001

For : SECRETED AND  
TRANSMEMBRANE  
POLYPEPTIDES AND NUCLEIC  
ACIDS ENCODING THE SAME

Examiner : Hamud, Fozia M

Group Art Unit 1647

CERTIFICATE OF EXPRESS MAILING

I hereby certify that this correspondence is being deposited with the United States Postal Service with sufficient postage as first class mail in an envelope addressed to Commissioner of Patents, Washington D.C. 20231 on:

(Date)

Commissioner of Patents  
P.O. Box 1450  
Alexandria, VA 22313-1450

DECLARATION OF AVI ASHKENAZI, Ph.D UNDER 37 C.F.R. § 1.132

I, Avi Ashkenazi, Ph.D. declare and say as follows: -

1. I am Director and Staff Scientist at the Molecular Oncology Department of Genentech, Inc., South San Francisco, CA 94080.
2. I joined Genentech in 1988 as a postdoctoral fellow. Since then, I have investigated a variety of cellular signal transduction mechanisms, including apoptosis, and have developed technologies to modulate such mechanisms as a means of therapeutic intervention in cancer and autoimmune disease. I am currently involved in the investigation of a series of secreted proteins over-expressed in tumors, with the aim to identify useful targets for the development of therapeutic antibodies for cancer treatment.
3. My scientific Curriculum Vitae, including my list of publications, is attached to and forms part of this Declaration (Exhibit A).
4. Gene amplification is a process in which chromosomes undergo changes to contain multiple copies of certain genes that normally exist as a single copy, and is an important factor in the pathophysiology of cancer. Amplification of certain genes (e.g., Myc or Her2/Neu)

gives cancer cells a growth or survival advantage relative to normal cells, and might also provide a mechanism of tumor cell resistance to chemotherapy or radiotherapy.

5. If gene amplification results in over-expression of the mRNA and the corresponding gene product, then it identifies that gene product as a promising target for cancer therapy, for example by the therapeutic antibody approach. Even in the absence of over-expression of the gene product, amplification of a cancer marker gene - as detected, for example, by the reverse transcriptase TaqMan<sup>®</sup> PCR or the fluorescence *in situ* hybridization (FISH) assays - is useful in the diagnosis or classification of cancer, or in predicting or monitoring the efficacy of cancer therapy. An increase in gene copy number can result not only from intrachromosomal changes but also from chromosomal aneuploidy. It is important to understand that detection of gene amplification can be used for cancer diagnosis even if the determination includes measurement of chromosomal aneuploidy. Indeed, as long as a significant difference relative to normal tissue is detected, it is irrelevant if the signal originates from an increase in the number of gene copies per chromosome and/or an abnormal number of chromosomes.

6. I understand that according to the Patent Office, absent data demonstrating that the increased copy number of a gene in certain types of cancer leads to increased expression of its product, gene amplification data are insufficient to provide substantial utility or well established utility for the gene product (the encoded polypeptide), or an antibody specifically binding the encoded polypeptide. However, even when amplification of a cancer marker gene does not result in significant over-expression of the corresponding gene product, this very absence of gene product over-expression still provides significant information for cancer diagnosis and treatment. Thus, if over-expression of the gene product does not parallel gene amplification in certain tumor types but does so in others, then parallel monitoring of gene amplification and gene product over-expression enables more accurate tumor classification and hence better determination of suitable therapy. In addition, absence of over-expression is crucial information for the practicing clinician. If a gene is amplified but the corresponding gene product is not over-expressed, the clinician accordingly will decide not to treat a patient with agents that target that gene product.

7. I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information or belief are believed to be true, and further that these statements were made with the knowledge that willful false statements and the like so

made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful statements may jeopardize the validity of the application or any patent issued thereon.

By: Avi Ashkenazi  
Avi Ashkenazi, Ph.D.

Date: 9/15/03

## **CURRICULUM VITAE**

**Avi Ashkenazi**

July 2003

### **Personal:**

Date of birth: 29 November, 1956  
Address: 1456 Tarrytown Street, San Mateo, CA 94402  
Phone: (650) 578-9199 (home); (650) 225-1853 (office)  
Fax: (650) 225-6443 (office)  
Email: aa@gene.com

### **Education:**

1983: B.S. in Biochemistry, with honors, Hebrew University, Israel  
1986: Ph.D. in Biochemistry, Hebrew University, Israel

### **Employment:**

1983-1986: Teaching assistant, undergraduate level course in Biochemistry  
1985-1986: Teaching assistant, graduate level course on Signal Transduction  
1986 - 1988: Postdoctoral fellow, Hormone Research Dept., UCSF, and  
Developmental Biology Dept., Genentech, Inc., with J. Ramachandran  
1988 - 1989: Postdoctoral fellow, Molecular Biology Dept., Genentech, Inc.,  
with D. Capon  
1989 - 1993: Scientist, Molecular Biology Dept., Genentech, Inc.  
1994 - 1996: Senior Scientist, Molecular Oncology Dept., Genentech, Inc.  
1996-1997: Senior Scientist and Interim director, Molecular Oncology Dept.,  
Genentech, Inc.  
1997-1990: Senior Scientist and preclinical project team leader, Genentech, Inc.  
1999 - 2002: Staff Scientist in Molecular Oncology, Genentech, Inc.  
2002-present: Staff Scientist and Director in Molecular Oncology, Genentech, Inc.

### **Awards:**

1988: First prize, The Boehringer Ingelheim Award

### Editorial:

Editorial Board Member: Current Biology

Associate Editor, Clinical Cancer Research.

Associate Editor, Cancer Biology and Therapy.

### Refereed papers:

1. Gertler, A., Ashkenazi, A., and Madar, Z. Binding sites for human growth hormone and ovine and bovine prolactins in the mammary gland and liver of the lactating cow. *Mol. Cell. Endocrinol.* 34, 51-57 (1984).
2. Gertler, A., Shamay, A., Cohen, N., Ashkenazi, A., Friesen, H., Levanon, A., Gorecki, M., Aviv, H., Hadari, D., and Vogel, T. Inhibition of lactogenic activities of ovine prolactin and human growth hormone (hGH) by a novel form of a modified recombinant hGH. *Endocrinology* 118, 720-726 (1986).
3. Ashkenazi, A., Madar, Z., and Gertler, A. Partial purification and characterization of bovine mammary gland prolactin receptor. *Mol. Cell. Endocrinol.* 50, 79-87 (1987).
4. Ashkenazi, A., Pines, M., and Gertler, A. Down-regulation of lactogenic hormone receptors in Nb2 lymphoma cells by cholera toxin. *Biochemistry Internatl.* 14, 1065-1072 (1987).
5. Ashkenazi, A., Cohen, R., and Gertler, A. Characterization of lactogen receptors in lactogenic hormone-dependent and independent Nb2 lymphoma cell lines. *FEBS Lett.* 210, 51-55 (1987).
6. Ashkenazi, A., Vogel, T., Barash, I., Hadari, D., Levanon, A., Gorecki, M., and Gertler, A. Comparative study on in vitro and in vivo modulation of lactogenic and somatotrophic receptors by native human growth hormone and its modified recombinant analog. *Endocrinology* 121, 414-419 (1987).
7. Peralta, E., Winslow, J., Peterson, G., Smith, D., Ashkenazi, A., Ramachandran, J., Schimerlik, M., and Capon, D. Primary structure and biochemical properties of an M2 muscarinic receptor. *Science* 236, 600-605 (1987).
8. Peralta, E., Ashkenazi, A., Winslow, J., Smith, D., Ramachandran, J., and Capon, D. J. Distinct primary structures, ligand-binding properties and tissue-specific expression of four human muscarinic acetylcholine receptors. *EMBO J.* 6, 3923-3929 (1987).
9. Ashkenazi, A., Winslow, J., Peralta, E., Peterson, G., Schimerlik, M., Capon, D., and Ramachandran, J. An M2 muscarinic receptor subtype coupled to both adenylyl cyclase and phosphoinositide turnover. *Science* 238, 672-675 (1987).

10. Pines, M., Ashkenazi, A., Cohen-Chapnik, N., Binder, L., and Gertler, A. Inhibition of the proliferation of Nb2 lymphoma cells by femtomolar concentrations of cholera toxin and partial reversal of the effect by 12-o-tetradecanoyl-phorbol-13-acetate. *J. Cell. Biochem.* 37, 119-129 (1988).
11. Peralta, E., Ashkenazi, A., Winslow, J., Ramachandran, J., and Capon, D. Differential regulation of PI hydrolysis and adenylyl cyclase by muscarinic receptor subtypes. *Nature* 334, 434-437 (1988).
12. Ashkenazi, A., Peralta, E., Winslow, J., Ramachandran, J., and Capon, D. Functionally distinct G proteins couple different receptors to PI hydrolysis in the same cell. *Cell* 56, 487-493 (1989).
13. Ashkenazi, A., Ramachandran, J., and Capon, D. Acetylcholine analogue stimulates DNA synthesis in brain-derived cells via specific muscarinic acetylcholine receptor subtypes. *Nature* 340, 146-150 (1989).
14. Lammare, D., Ashkenazi, A., Fleury, S., Smith, D., Sekaly, R., and Capon, D. The MHC-binding and gp120-binding domains of CD4 are distinct and separable. *Science* 245, 743-745 (1989).
15. Ashkenazi, A., Presta, L., Marsters, S., Camerato, T., Rosenthal, K., Fendly, B., and Capon, D. Mapping the CD4 binding site for human immunodeficiency virus type 1 by alanine-scanning mutagenesis. *Proc. Natl. Acad. Sci. USA.* 87, 7150-7154 (1990).
16. Chamow, S., Peers, D., Byrn, R., Mulkerrin, M., Harris, R., Wang, W., Bjorkman, P., Capon, D., and Ashkenazi, A. Enzymatic cleavage of a CD4 immunoadhesin generates crystallizable, biologically active Fd-like fragments. *Biochemistry* 29, 9885-9891 (1990).
17. Ashkenazi, A., Smith, D., Marsters, S., Riddle, L., Gregory, T., Ho, D., and Capon, D. Resistance of primary isolates of human immunodeficiency virus type 1 to soluble CD4 is independent of CD4-rgp120 binding affinity. *Proc. Natl. Acad. Sci. USA.* 88, 7056-7060 (1991).
18. Ashkenazi, A., Marsters, S., Capon, D., Chamow, S., Figari, I., Pennica, D., Goeddel, D., Palladino, M., and Smith, D. Protection against endotoxic shock by a tumor necrosis factor receptor immunoadhesin. *Proc. Natl. Acad. Sci. USA.* 88, 10535-10539 (1991).
19. Moore, J., McKeating, J., Huang, Y., Ashkenazi, A., and Ho, D. Virions of primary HIV-1 isolates resistant to sCD4 neutralization differ in sCD4 affinity and glycoprotein gp120 retention from sCD4-sensitive isolates. *J. Virol.* 66, 235-243 (1992).

20. Jin, H., Oksenberg, D., Ashkenazi, A., Peroutka, S., Duncan, A., Rozmahel, R., Yang, Y., Mengod, G., Palacios, J., and O'Dowd, B. Characterization of the human 5-hydroxytryptamine<sub>1B</sub> receptor. *J. Biol. Chem.* 267, 5735-5738 (1992).
21. Marsters, A., Frutkin, A., Simpson, N., Fendly, B. and Ashkenazi, A. Identification of cysteine-rich domains of the type 1 tumor necrosis receptor involved in ligand binding. *J. Biol. Chem.* 267, 5747-5750 (1992).
22. Chamow, S., Kogan, T., Peers, D., Hastings, R., Byrn, R., and Ashkenazi, A. Conjugation of sCD4 without loss of biological activity via a novel carbohydrate-directed cross-linking reagent. *J. Biol. Chem.* 267, 15916-15922 (1992).
23. Oksenberg, D., Marsters, A., O'Dowd, B., Jin, H., Havlik, S., Peroutka, S., and Ashkenazi, A. A single amino-acid difference confers major pharmacologic variation between human and rodent 5-HT<sub>1B</sub> receptors. *Nature* 360, 161-163 (1992).
24. Haak-Frendscho, M., Marsters, S., Chamow, S., Peers, D., Simpson, N., and Ashkenazi, A. Inhibition of interferon  $\gamma$  by an interferon  $\gamma$  receptor immunoadhesin. *Immunology* 79, 594-599 (1993).
25. Penica, D., Lam, V., Weber, R., Kohr, W., Basa, L., Spellman, M., Ashkenazi, A., Shire, S., and Goeddel, D. Biochemical characterization of the extracellular domain of the 75-kd tumor necrosis factor receptor. *Biochemistry* 32, 3131-3138. (1993).
26. Barford, L., Zheng, Y., Kuang, W., Hart, M., Evans, T., Cerione, R., and Ashkenazi, A. Cloning and expression of a human CDC42 GTPase Activating Protein reveals a functional SH3-binding domain. *J. Biol. Chem.* 268, 26059-26062 (1993).
27. Chamow, S., Zhang, D., Tan, X., Mhtre, S., Marsters, S., Peers, D., Byrn, R., Ashkenazi, A., and Yunghans, R. A humanized bispecific immunoadhesin-antibody that retargets CD3<sup>+</sup> effectors to kill HIV-1-infected cells. *J. Immunol.* 153, 4268-4280 (1994).
28. Means, R., Krantz, S., Luna, J., Marsters, S., and Ashkenazi, A. Inhibition of murine erythroid colony formation in vitro by iterferon  $\gamma$  and correction by interferon  $\gamma$  receptor immunoadhesin. *Blood* 83, 911-915 (1994).
29. Haak-Frendscho, M., Marsters, S., Mordenti, J., Gillet, N., Chen, S., and Ashkenazi, A. Inhibition of TNF by a TNF receptor immunoadhesin: comparison with an anti-TNF mAb. *J. Immunol.* 152, 1347-1353 (1994).

30. Chamow, S., Kogan, T., Venuti, M., Gadek, T., Peers, D., Mordenti, J., Shak, S., and Ashkenazi, A. Modification of CD4 immunoadhesin with monomethoxy-PEG aldehyde via reductive alkylation. *Bioconj. Chem.* 5, 133-140 (1994).
31. Jin, H., Yang, R., Marsters, S., Bunting, S., Wurm, F., Chamow, S., and Ashkenazi, A. Protection against rat endotoxic shock by p55 tumor necrosis factor (TNF) receptor immunoadhesin: comparison to anti-TNF monoclonal antibody. *J. Infect. Diseases* 170, 1323-1326 (1994).
32. Beck, J., Marsters, S., Harris, R., Ashkenazi, A., and Chamow, S. Generation of soluble interleukin-1 receptor from an immunoadhesin by specific cleavage. *Mol. Immunol.* 31, 1335-1344 (1994).
33. Pitti, B., Marsters, M., Haak-Frendscho, M., Osaka, G., Mordenti, J., Chamow, S., and Ashkenazi, A. Molecular and biological properties of an interleukin-1 receptor immunoadhesin. *Mol. Immunol.* 31, 1345-1351 (1994).
34. Oksenberg, D., Havlik, S., Peroutka, S., and Ashkenazi, A. The third intracellular loop of the 5-HT<sub>2</sub> receptor specifies effector coupling. *J. Neurochem.* 64, 1440-1447 (1995).
35. Bach, E., Szabo, S., Dighe, A., Ashkenazi, A., Aguet, M., Murphy, K., and Schreiber, R. Ligand-induced autoregulation of IFN- $\gamma$  receptor  $\beta$  chain expression in T helper cell subsets. *Science* 270, 1215-1218 (1995).
36. Jin, H., Yang, R., Marsters, S., Ashkenazi, A., Bunting, S., Marra, M., Scott, R., and Baker, J. Protection against endotoxic shock by bactericidal/permeability-increasing protein in rats. *J. Clin. Invest.* 95, 1947-1952 (1995).
37. Marsters, S., Penica, D., Bach, E., Schreiber, R., and Ashkenazi, A. Interferon  $\gamma$  signals via a high-affinity multisubunit receptor complex that contains two types of polypeptide chain. *Proc. Natl. Acad. Sci. USA.* 92, 5401-5405 (1995).
38. Van Zee, K., Moldawer, L., Oldenburg, H., Thompson, W., Stackpole, S., Montegut, W., Rogy, M., Meschter, C., Gallati, H., Schiller, C., Richter, W., Loetcher, H., Ashkenazi, A., Chamow, S., Wurm, F., Calvano, S., Lowry, S., and Lesslauer, W. Protection against lethal E. coli bacteremia in baboons by pretreatment with a 55-kDa TNF receptor-Ig fusion protein, Ro45-2081. *J. Immunol.* 156, 2221-2230 (1996).
39. Pitti, R., Marsters, S., Ruppert, S., Donahue, C., Moore, A., and Ashkenazi, A. Induction of apoptosis by Apo-2 Ligand, a new member of the tumor necrosis factor cytokine family. *J. Biol. Chem.* 271, 12687-12690 (1996).

40. Marsters, S., Pitti, R., Donahue, C., Rupert, S., Bauer, K., and Ashkenazi, A. Activation of apoptosis by Apo-2 ligand is independent of FADD but blocked by CrmA. *Curr. Biol.* 6, 1669-1676 (1996).
41. Marsters, S., Skubatch, M., Gray, C., and Ashkenazi, A. Herpesvirus entry mediator, a novel member of the tumor necrosis factor receptor family, activates the NF- $\kappa$ B and AP-1 transcription factors. *J. Biol. Chem.* 272, 14029-14032 (1997).
42. Sheridan, J., Marsters, S., Pitti, R., Gurney, A., Skubatch, M., Baldwin, D., Ramakrishnan, L., Gray, C., Baker, K., Wood, W.I., Goddard, A., Godowski, P., and Ashkenazi, A. Control of TRAIL-induced apoptosis by a family of signaling and decoy receptors. *Science* 277, 818-821 (1997).
43. Marsters, S., Sheridan, J., Pitti, R., Gurney, A., Skubatch, M., Baldwin, D., Huang, A., Yuan, J., Goddard, A., Godowski, P., and Ashkenazi, A. A novel receptor for Apo2L/TRAIL contains a truncated death domain. *Curr. Biol.* 7, 1003-1006 (1997).
44. Marsters, A., Sheridan, J., Pitti, R., Brush, J., Goddard, A., and Ashkenazi, A. Identification of a ligand for the death-domain-containing receptor Apo3. *Curr. Biol.* 8, 525-528 (1998).
45. Rieger, J., Naumann, U., Glaser, T., Ashkenazi, A., and Weller, M. Apo2 ligand: a novel weapon against malignant glioma? *FEBS Lett.* 427, 124-128 (1998).
46. Pender, S., Fell, J., Chamow, S., Ashkenazi, A., and MacDonald, T. A p55 TNF receptor immunoadhesin prevents T cell mediated intestinal injury by inhibiting matrix metalloproteinase production. *J. Immunol.* 160, 4098-4103 (1998).
47. Pitti, R., Marsters, S., Lawrence, D., Roy, Kischkel, F., M., Dowd, P., Huang, A., Donahue, C., Sherwood, S., Baldwin, D., Godowski, P., Wood, W., Gurney, A., Hillan, K., Cohen, R., Goddard, A., Botstein, D., and Ashkenazi, A. Genomic amplification of a decoy receptor for Fas ligand in lung and colon cancer. *Nature* 396, 699-703 (1998).
48. Mori, S., Marakami-Mori, K., Nakamura, S., Ashkenazi, A., and Bonavida, B. Sensitization of AIDS Kaposi's sarcoma cells to Apo-2 ligand-induced apoptosis by actinomycin D. *J. Immunol.* 162, 5616-5623 (1999).
49. Gurney, A. Marsters, S., Huang, A., Pitti, R., Mark, M., Baldwin, D., Gray, A., Dowd, P., Brush, J., Heldens, S., Schow, P., Goddard, A., Wood, W., Baker, K., Godowski, P., and Ashkenazi, A. Identification of a new member of the tumor necrosis factor family and its receptor, a human ortholog of mouse GITR. *Curr. Biol.* 9, 215-218 (1999).

50. Ashkenazi, A., Pai, R., Fong, s., Leung, S., Lawrence, D., Marsters, S., Blackie, C., Chang, L., McMurtrey, A., Hebert, A., DeForge, L., Khoumenis, I., Lewis, D., Harris, L., Bussiere, J., Koeppen, H., Shahrokh, Z., and Schwall, R. Safety and anti-tumor activity of recombinant soluble Apo2 ligand. *J. Clin. Invest.* 104, 155-162 (1999).
51. Chuntharapai, A., Gibbs, V., Lu, J., Ow, A., Marsters, S., Ashkenazi, A., De Vos, A., Kim, K.J. Determination of residues involved in ligand binding and signal transmissiion in the human IFN- $\alpha$  receptor 2. *J. Immunol.* 163, 766-773 (1999).
52. Johnsen, A.-C., Haux, J., Steinkjer, B., Nonstad, U., Egeberg, K., Sundan, A., Ashkenazi, A., and Espevik, T. Regulation of Apo2L/TRAIL expression in NK cells – involvement in NK cell-mediated cytotoxicity. *Cytokine* 11, 664-672 (1999).
53. Roth, W., Isenmann, S., Naumann, U., Kugler, S., Bahr, M., Dichgans, J., Ashkenazi, A., and Weller, M. Eradication of intracranial human malignant glioma xenografts by Apo2L/TRAIL. *Biochem. Biophys. Res. Commun.* 265, 479-483 (1999).
54. Hymowitz, S.G., Christinger, H.W., Fuh, G., Ultsch, M., O'Connell, M., Kelley, R.F., Ashkenazi, A. and de Vos, A.M. Triggering Cell Death: The Crystal Structure of Apo2L/TRAIL in a Complex with Death Receptor 5. *Molec. Cell* 4, 563-571 (1999).
55. Hymowitz, S.G., O'Connel, M.P., Utsch, M.H., Hurst, A., Totpal, K., Ashkenazi, A., de Vos, A.M., Kelley, R.F. A unique zinc-binding site revealed by a high-resolution X-ray structure of homotrimeric Apo2L/TRAIL. *Biochemistry* 39, 633-640 (2000).
56. Zhou, Q., Fukushima, P., DeGraff, W., Mitchell, J.B., Stetler-Stevenson, M., Ashkenazi, A., and Steeg, P.S. Radiation and the Apo2L/TRAIL apoptotic pathway preferentially inhibit the colonization of premalignant human breast cancer cells overexpressing cyclin D1. *Cancer Res.* 60, 2611-2615 (2000).
57. Kischkel, F.C., Lawrence, D. A., Chuntharapai, A., Schow, P., Kim, J., and Ashkenazi, A. Apo2L/TRAIL-dependent recruitment of endogenous FADD and Caspase-8 to death receptors 4 and 5. *Immunity* 12, 611-620 (2000).
58. Yan, M., Marsters, S.A., Grewal, I.S., Wang, H., \*Ashkenazi, A., and \*Dixit, V.M. Identification of a receptor for BlyS demonstrates a crucial role in humoral immunity. *Nature Immunol.* 1, 37-41 (2000).

59. Marsters, S.A., Yan, M., Pitti, R.M., Haas, P.E., Dixit, V.M., and Ashkenazi, A. Interaction of the TNF homologues BlyS and APRIL with the TNF receptor homologues BCMA and TACI. *Curr. Biol.* 10, 785-788 (2000).
60. Kischkel, F.C., and Ashkenazi, A. Combining enhanced metabolic labeling with immunoblotting to detect interactions of endogenous cellular proteins. *Biotechniques* 29, 506-512 (2000).
61. Lawrence, D., Shahrokh, Z., Marsters, S., Achilles, K., Shih, D. Mounho, B., Hillan, K., Totpal, K. DeForge, L., Schow, P., Hooley, J., Sherwood, S., Pai, R., Leung, S., Khan, L., Gliniak, B., Bussiere, J., Smith, C., Strom, S., Kelley, S., Fox, J., Thomas, D., and Ashkenazi, A. Differential hepatocyte toxicity of recombinant Apo2L/TRAIL versions. *Nature Med.* 7, 383-385 (2001).
62. Chuntharapai, A., Dodge, K., Grimmer, K., Schroeder, K., Marsters, S.A., Koeppen, H., Ashkenazi, A., and Kim, K.J. Isotype-dependent inhibition of tumor growth in vivo by monoclonal antibodies to death receptor 4. *J. Immunol.* 166, 4891-4898 (2001).
63. Pollack, I.F., Erff, M., and Ashkenazi, A. Direct stimulation of apoptotic signaling by soluble Apo2L/tumor necrosis factor-related apoptosis-inducing ligand leads to selective killing of glioma cells. *Clin. Cancer Res.* 7, 1362-1369 (2001).
64. Wang, H., Marsters, S.A., Baker, T., Chan, B., Lee, W.P., Fu, L., Tumas, D., Yan, M., Dixit, V.M., \*Ashkenazi, A., and \*Grewal, I.S. TACI-ligand interactions are required for T cell activation and collagen-induced arthritis in mice. *Nature Immunol.* 2, 632-637 (2001).
65. Kischkel, F.C., Lawrence, D. A., Tinel, A., Virmani, A., Schow, P., Gazdar, A., Blenis, J., Arnott, D., and Ashkenazi, A. Death receptor recruitment of endogenous caspase-10 and apoptosis initiation in the absence of caspase-8. *J. Biol. Chem.* 276, 46639-46646 (2001).
66. LeBlanc, H., Lawrence, D.A., Varfolomeev, E., Totpal, K., Morlan, J., Schow, P., Fong, S., Schwall, R., Sinicropi, D., and Ashkenazi, A. Tumor cell resistance to death receptor induced apoptosis through mutational inactivation of the proapoptotic Bcl-2 homolog Bax. *Nature Med.* 8, 274-281 (2002).
67. Miller, K., Meng, G., Liu, J., Hurst, A., Hsei, V., Wong, W-L., Ekert, R., Lawrence, D., Sherwood, S., DeForge, L., Gaudreault, G., Keller, G., Sliwkowski, M., Ashkenazi, A., and Presta, L. Design, Construction, and analyses of multivalent antibodies. *J. Immunol.* 170, 4854-4861 (2003).

68. Varfolomeev, E., Kischkel, F., Martin, F., Wanh, H., Lawrence, D., Olsson, C., Tom, L., Erickson, S., French, D., Schow, P., Grewal, I. and Ashkenazi, A. Immune system development in APRIL knockout mice. Submitted.

**Review articles:**

1. Ashkenazi, A., Peralta, E., Winslow, J., Ramachandran, J., and Capon, D., J. Functional role of muscarinic acetylcholine receptor subtype diversity. *Cold Spring Harbor Symposium on Quantitative Biology*. LIII, 263-272 (1988).
2. Ashkenazi, A., Peralta, E., Winslow, J., Ramachandran, J., and Capon, D. Functional diversity of muscarinic receptor subtypes in cellular signal transduction and growth. *Trends Pharmacol. Sci.* Dec Supplement, 12-21 (1989).
3. Chamow, S., Duliege, A., Ammann, A., Kahn, J., Allen, D., Eichberg, J., Byrn, R., Capon, D., Ward, R., and Ashkenazi, A. CD4 immunoadhesins in anti-HIV therapy: new developments. *Int. J. Cancer* Supplement 7, 69-72 (1992).
4. Ashkenazi, A., Capon, and D. Ward, R. Immunoadhesins. *Int. Rev. Immunol.* 10, 217-225 (1993).
5. Ashkenazi, A., and Peralta, E. Muscarinic Receptors. In *Handbook of Receptors and Channels*. (S. Peroutka, ed.), CRC Press, Boca Raton, Vol. I, p. 1-27, (1994).
6. Krantz, S. B., Means, R. T., Jr., Lina, J., Marsters, S. A., and Ashkenazi, A. Inhibition of erythroid colony formation in vitro by gamma interferon. In *Molecular Biology of Hematopoiesis* (N. Abraham, R. Shadduck, A. Levine F. Takaku, eds.) Intercept Ltd. Paris, Vol. 3, p. 135-147 (1994).
7. Ashkenazi, A. Cytokine neutralization as a potential therapeutic approach for SIRS and shock. *J. Biotechnology in Healthcare* 1, 197-206 (1994).
8. Ashkenazi, A., and Chamow, S. M. Immunoadhesins: an alternative to human monoclonal antibodies. *Immunomethods: A companion to Methods in Enzymology* 8, 104-115 (1995).
9. Chamow, S., and Ashkenazi, A. Immunoadhesins: Principles and Applications. *Trends Biotech.* 14, 52-60 (1996).
10. Ashkenazi, A., and Chamow, S. M. Immunoadhesins as research tools and therapeutic agents. *Curr. Opin. Immunol.* 9, 195-200 (1997).
11. Ashkenazi, A., and Dixit, V. Death receptors: signaling and modulation. *Science* 281, 1305-1308 (1998).
12. Ashkenazi, A., and Dixit, V. Apoptosis control by death and decoy receptors. *Curr. Opin. Cell. Biol.* 11, 255-260 (1999).

13. Ashkenazi, A. Chapters on Apo2L/TRAIL; DR4, DR5, DcR1, DcR2; and DcR3. Online Cytokine Handbook ([www.apnet.com/cytokinereference/](http://www.apnet.com/cytokinereference/)).
14. Ashkenazi, A. Targeting death and decoy receptors of the tumor necrosis factor superfamily. *Nature Rev. Cancer* 2, 420-430 (2002).
15. LeBlanc, H. and Ashkenazi, A. Apoptosis signaling by Apo2L/TRAIL. *Cell Death and Differentiation* 10, 66-75 (2003).
16. Almasan, A. and Ashkenazi, A. Apo2L/TRAIL: apoptosis signaling, biology, and potential for cancer therapy. *Cytokine and Growth Factor Reviews* 14, 337-348 (2003).

**Book:**

Antibody Fusion Proteins (Chamow, S., and Ashkenazi, A., eds., John Wiley and Sons Inc.) (1999).

**Talks:**

1. Resistance of primary HIV isolates to CD4 is independent of CD4-gp120 binding affinity. UCSD Symposium, HIV Disease: Pathogenesis and Therapy. Greenelefe, FL, March 1991.
2. Use of immuno-hybrids to extend the half-life of receptors. IBC conference on Biopharmaceutical Half-life Extension. New Orleans, LA, June 1992.
3. Results with TNF receptor Immunoadhesins for the Treatment of Sepsis. IBC conference on Endotoxemia and Sepsis. Philadelphia, PA, June 1992.
4. Immunoadhesins: an alternative to human antibodies. IBC conference on Antibody Engineering. San Diego, CA, December 1993.
5. Tumor necrosis factor receptor: a potential therapeutic for human septic shock. American Society for Microbiology Meeting, Atlanta, GA, May 1993.
6. Protective efficacy of TNF receptor immunoadhesin vs anti-TNF monoclonal antibody in a rat model for endotoxic shock. 5th International Congress on TNF. Asilomar, CA, May 1994.
7. Interferon- $\gamma$  signals via a multisubunit receptor complex that contains two types of polypeptide chain. American Association of Immunologists Conference. San Francisco, CA, July 1995.
8. Immunoadhesins: Principles and Applications. Gordon Research Conference on Drug Delivery in Biology and Medicine. Ventura, CA, February 1996.

9. Apo-2 Ligand, a new member of the TNF family that induces apoptosis in tumor cells. Cambridge Symposium on TNF and Related Cytokines in Treatment of Cancer. Hilton-Head, NC, March 1996.
10. Induction of apoptosis by Apo2 Ligand. American Society for Biochemistry and Molecular Biology, Symposium on Growth Factors and Cytokine Receptors. New Orleans, LA, June, 1996.
11. Apo2 ligand, an extracellular trigger of apoptosis. 2nd Clontech Symposium, Palo Alto, CA, October 1996.
12. Regulation of apoptosis by members of the TNF ligand and receptor families. Stanford University School of Medicine, Palo Alto, CA, December 1996.
13. Apo-3: a novel receptor that regulates cell death and inflammation. 4th International Congress on Immune Consequences of Trauma, Shock, and Sepsis. Munich, Germany, March 1997.
14. New members of the TNF ligand and receptor families that regulate apoptosis, inflammation, and immunity. UCLA School of Medicine, LA, CA, March 1997.
15. Immunoadhesins: an alternative to monoclonal antibodies. 5th World Conference on Bispecific Antibodies. Volendam, Holland, June 1997.
16. Control of Apo2L signaling. Cold Spring Harbor Laboratory Symposium on Programmed Cell Death. Cold Spring Harbor, New York. September, 1997.
17. Chairman and speaker, Apoptosis Signaling session. IBC's 4th Annual Conference on Apoptosis. San Diego, CA., October 1997.
18. Control of Apo2L signaling by death and decoy receptors. American Association for the Advancement of Science. Philadelphia, PA, February 1998.
19. Apo2 ligand and its receptors. American Society of Immunologists. San Francisco, CA, April 1998.
20. Death receptors and ligands. 7th International TNF Congress. Cape Cod, MA, May 1998.
21. Apo2L as a potential therapeutic for cancer. UCLA School of Medicine. LA, CA, June 1998.
22. Apo2L as a potential therapeutic for cancer. Gordon Research Conference on Cancer Chemotherapy. New London, NH, July 1998.
23. Control of apoptosis by Apo2L. Endocrine Society Conference, Stevenson, WA, August 1998.
24. Control of apoptosis by Apo2L. International Cytokine Society Conference, Jerusalem, Israel, October 1998.

25. Apoptosis control by death and decoy receptors. American Association for Cancer Research Conference, Whistler, BC, Canada, March 1999.
26. Apoptosis control by death and decoy receptors. American Society for Biochemistry and Molecular Biology Conference, San Francisco, CA, May 1999.
27. Apoptosis control by death and decoy receptors. Gordon Research Conference on Apoptosis, New London, NH, June 1999.
28. Apoptosis control by death and decoy receptors. Arthritis Foundation Research Conference, Alexandria GA, Aug 1999.
29. Safety and anti-tumor activity of recombinant soluble Apo2L/TRAIL. Cold Spring Harbor Laboratory Symposium on Programmed Cell Death. Cold Spring Harbor, NY, September 1999.
30. The Apo2L/TRAIL system: therapeutic potential. American Association for Cancer Research, Lake Tahoe, NV, Feb 2000.
31. Apoptosis and cancer therapy. Stanford University School of Medicine, Stanford, CA, Mar 2000.
32. Apoptosis and cancer therapy. University of Pennsylvania School of Medicine, Philadelphia, PA, Apr 2000.
33. Apoptosis signaling by Apo2L/TRAIL. International Congress on TNF. Trondheim, Norway, May 2000.
34. The Apo2L/TRAIL system: therapeutic potential. Cap-CURE summit meeting. Santa Monica, CA, June 2000.
35. The Apo2L/TRAIL system: therapeutic potential. MD Anderson Cancer Center. Houston, TX, June 2000.
36. Apoptosis signaling by Apo2L/TRAIL. The Protein Society, 14<sup>th</sup> Symposium. San Diego, CA, August 2000.
37. Anti-tumor activity of Apo2L/TRAIL. AAPS annual meeting. Indianapolis, IN Aug 2000.
38. Apoptosis signaling and anti-cancer potential of Apo2L/TRAIL. Cancer Research Institute, UC San Francisco, CA, September 2000.
39. Apoptosis signaling by Apo2L/TRAIL. Kenote address, TNF family Minisymposium, NIH. Bethesda, MD, September 2000.
40. Death receptors: signaling and modulation. Keystone symposium on the Molecular basis of cancer. Taos, NM, Jan 2001.
41. Preclinical studies of Apo2L/TRAIL in cancer. Symposium on Targeted therapies in the treatment of lung cancer. Aspen, CO, Jan 2001.

42. Apoptosis signaling by Apo2L/TRAIL. Weizmann Institute of Science, Rehovot, Israel, March 2001.
43. Apo2L/TRAIL: Apoptosis signaling and potential for cancer therapy. Weizmann Institute of Science, Rehovot, Israel, March 2001.
44. Targeting death receptors in cancer with Apo2L/TRAIL. Cell Death and Disease conference, North Falmouth, MA, Jun 2001.
45. Targeting death receptors in cancer with Apo2L/TRAIL. Biotechnology Organization conference, San Diego, CA, Jun 2001.
46. Apo2L/TRAIL signaling and apoptosis resistance mechanisms. Gordon Research Conference on Apoptosis, Oxford, UK, July 2001.
47. Apo2L/TRAIL signaling and apoptosis resistance mechanisms. Cleveland Clinic Foundation, Cleveland, OH, Oct 2001.
48. Apoptosis signaling by death receptors: overview. International Society for Interferon and Cytokine Research conference, Cleveland, OH, Oct 2001.
49. Apoptosis signaling by death receptors. American Society of Nephrology Conference. San Francisco, CA, Oct 2001.
50. Targeting death receptors in cancer. Apoptosis: commercial opportunities. San Diego, CA, Apr 2002.
51. Apo2L/TRAIL signaling and apoptosis resistance mechanisms. Kimmel Cancer Research Center, Johns Hopkins University, Baltimore MD. May 2002.
52. Apoptosis control by Apo2L/TRAIL. (Keynote Address) University of Alabama Cancer Center Retreat, Birmingham, Ab. October 2002.
53. Apoptosis signaling by Apo2L/TRAIL. (Session co-chair) TNF international conference. San Diego, CA. October 2002.
54. Apoptosis signaling by Apo2L/TRAIL. Swiss Institute for Cancer Research (ISREC). Lausanne, Switzerland. Jan 2003.
55. Apoptosis induction with Apo2L/TRAIL. Conference on New Targets and Innovative Strategies in Cancer Treatment. Monte Carlo. February 2003.
56. Apoptosis signaling by Apo2L/TRAIL. Hermelin Brain Tumor Center Symposium on Apoptosis. Detroit, MI. April 2003.
57. Targeting apoptosis through death receptors. Sixth Annual Conference on Targeted Therapies in the Treatment of Breast Cancer. Kona, Hawaii. July 2003.
58. Targeting apoptosis through death receptors. Second International Conference on Targeted Cancer Therapy. Washington, DC. Aug 2003.

#### **Issued Patents:**

1. Ashkenazi, A., Chamow, S. and Kogan, T. Carbohydrate-directed crosslinking reagents. US patent 5,329,028 (Jul 12, 1994).
2. Ashkenazi, A., Chamow, S. and Kogan, T. Carbohydrate-directed crosslinking reagents. US patent 5,605,791 (Feb 25, 1997).
3. Ashkenazi, A., Chamow, S. and Kogan, T. Carbohydrate-directed crosslinking reagents. US patent 5,889,155 (Jul 27, 1999).
4. Ashkenazi, A., APO-2 Ligand. US patent 6,030,945 (Feb 29, 2000).
5. Ashkenazi, A., Chuntharapai, A., Kim, J., APO-2 ligand antibodies. US patent 6,046,048 (Apr 4, 2000).
6. Ashkenazi, A., Chamow, S. and Kogan, T. Carbohydrate-directed crosslinking reagents. US patent 6,124,435 (Sep 26, 2000).
7. Ashkenazi, A., Chuntharapai, A., Kim, J., Method for making monoclonal and cross-reactive antibodies. US patent 6,252,050 (Jun 26, 2001).
8. Ashkenazi, A. APO-2 Receptor. US patent 6,342,369 (Jan 29, 2002).
9. Ashkenazi, A. Fong, S., Goddard, A., Gurney, A., Napier, M., Tumas, D., Wood, W. A-33 polypeptides. US patent 6,410,708 (Jun 25, 2002).
10. Ashkenazi, A. APO-3 Receptor. US patent 6,462,176 B1 (Oct 8, 2002).
11. Ashkenazi, A. APO-2LI and APO-3 polypeptide antibodies. US patent 6,469,144 B1 (Oct 22, 2002).
12. Ashkenazi, A., Chamow, S. and Kogan, T. Carbohydrate-directed crosslinking reagents. US patent 6,582,928B1 (Jun 24, 2003).



# TECHNICAL UPDATE

FROM YOUR LABORATORY SERVICES PROVIDER

## HER-2/neu Breast Cancer Predictive Testing

*Julie Sanford Hanna, Ph.D. and Dan Mornin, M.D.*

EACH YEAR, OVER 182,000 WOMEN in the United States are diagnosed with breast cancer, and approximately 45,000 die of the disease.<sup>1</sup> Incidence appears to be increasing in the United States at a rate of roughly 2% per year. The reasons for the increase are unclear, but non-genetic risk factors appear to play a large role.<sup>2</sup>

Five-year survival rates range from approximately 65%-85%, depending on demographic group, with a significant percentage of women experiencing recurrence of their cancer within 10 years of diagnosis. One of the factors most predictive for recurrence once a diagnosis of breast cancer has been made is the number of axillary lymph nodes to which tumor has metastasized. Most node-positive women are given adjuvant therapy, which increases their survival. However, 20%-30% of patients without axillary node involvement also develop recurrent disease, and the difficulty lies in how to identify this high-risk subset of patients. These patients could benefit from increased surveillance, early intervention, and treatment.

Prognostic markers currently used in breast cancer recurrence prediction include tumor size, histological grade, steroid hormone receptor status, DNA ploidy, proliferative index, and cathepsin D status. Expression of growth factor receptors and over-expression of the HER-2/neu oncogene have also been identified as having value regarding treatment regimen and prognosis.

HER-2/neu (also known as c-erbB2) is an oncogene that encodes a transmembrane glycoprotein that is homologous to, but distinct from, the epidermal growth factor receptor. Numerous studies have indicated that high levels of expression of this protein are associated with rapid tumor growth, certain forms of therapy resistance, and shorter disease-free survival. The gene has been shown to be amplified and/or overexpressed in 10%-30% of invasive breast cancers and in 40%-60% of intraductal breast carcinoma.<sup>3</sup>

There are two distinct FDA-approved methods by which HER-2/neu status can be evaluated: immunohistochemistry (IHC, HercepTest™) and FISH (fluorescent in situ hybridization, PathVysion™ Kit). Both methods can be performed on archived and current specimens. The first method allows visual assessment of the amount of HER-2/neu protein present on the cell membrane. The latter method allows direct quantification of the level of gene amplification present in the tumor, enabling differentiation between low- versus high-amplification. At least one study has demonstrated a difference in

recurrence risk in women younger than 40 years of age for low- versus high-amplified tumors (54.5% compared to 85.7%); this is compared to a recurrence rate of 16.7% for patients with no HER-2/neu gene amplification.<sup>4</sup> HER-2/neu status may be particularly important to establish in women with small ( $\leq 1$  cm) tumor size.

The choice of methodology for determination of HER-2/neu status depends in part on the clinical setting. FDA approval for the Vysis FISH test was granted based on clinical trials involving 1549 node-positive patients. Patients received one of three different treatments consisting of different doses of cyclophosphamide, Adriamycin, and 5-fluorouracil (CAF). The study showed that patients with amplified HER-2/neu benefited from treatment with higher doses of adriamycin-based therapy, while those with normal HER-2/neu levels did not. The study therefore identified a sub-set of women, who because they did not benefit from more aggressive treatment, did not need to be exposed to the associated side effects. In addition, other evidence indicates that HER-2/neu amplification in node-negative patients can be used as an independent prognostic indicator for early recurrence, recurrent disease at any time and disease-related death.<sup>5</sup> Demonstration of HER-2/neu gene amplification by FISH has also been shown to be of value in predicting response to chemotherapy in stage-2 breast cancer patients.

Selection of patients for Herceptin® (Trastuzumab) monoclonal antibody therapy, however, is based upon demonstration of HER-2/neu protein overexpression using HercepTest™. Studies using Herceptin® in patients with metastatic breast cancer show an increase in time to disease progression, increased response rate to chemotherapeutic agents and a small increase in overall survival rate. The FISH assays have not yet been approved for this purpose, and studies looking at response to Herceptin® in patients with or without gene amplification status determined by FISH are in progress.

In general, FISH and IHC results correlate well. However, subsets of tumors are found which show discordant results; i.e., protein overexpression without gene amplification or lack of protein overexpression with gene amplification. The clinical significance of such results is unclear. Based on the above considerations, HER-2/neu testing at SHMC/PAML will utilize immunohistochemistry (HercepTest®) as a screen, followed by FISH in IHC-negative cases. Alternatively, either method may be ordered individually depending on the clinical setting or clinician preference.

## CPT code information

### HER-2/neu via IHC

88342 (including interpretive report)

### HER-2/neu via FISH

88271x2 Molecular cytogenetics, DNA probe, each

88274 Molecular cytogenetics, interphase in situ hybridization, analyze 25-99 cells

88291 Cytogenetics and molecular cytogenetics, interpretation and report

## Procedural Information

Immunohistochemistry is performed using the FDA-approved DAKO antibody kit, Herceptest<sup>®</sup>. The DAKO kit contains reagents required to complete a two-step immunohistochemical staining procedure for routinely processed, paraffin-embedded specimens. Following incubation with the primary rabbit antibody to human HER-2/neu protein, the kit employs a ready-to-use dextran-based visualization reagent. This reagent consists of both secondary goat anti-rabbit antibody molecules with horseradish peroxidase molecules linked to a common dextran polymer backbone, thus eliminating the need for sequential application of link antibody and peroxidase conjugated antibody. Enzymatic conversion of the subsequently added chromogen results in formation of visible reaction product at the antigen site. The specimen is then counterstained; a pathologist using light-microscopy interprets results.

FISH analysis at SHMC/PAML is performed using the FDA-approved PathVysion<sup>™</sup> HER-2/neu DNA probe kit, produced by Vysis, Inc. Formalin fixed, paraffin-embedded breast tissue is processed using routine histological methods, and then slides are treated to allow hybridization of DNA probes to the nuclei present in the tissue section. The PathVysion<sup>™</sup> kit contains two direct-labeled DNA probes, one specific for the alphoid repetitive DNA (CEP 17, spectrum orange) present at the chromosome 17 centromere and the second for the HER-2/neu oncogene located at 17q11.2-12 (spectrum green). Enumeration of the probes allows a ratio of the number of copies of chromosome 17 to the number of copies of HER-2/neu to be obtained; this enables quantification of low versus high amplification levels, and allows an estimate of the percentage of cells with HER-2/neu gene amplification. The clinically relevant distinction is whether the gene amplification is due to increased gene copy number on the two chromosome 17 homologues normally present or an increase in the number of chromosome 17s in the cells. In the majority of cases, ratio equivalents less than 2.0 are indicative of a normal/negative result, ratios of 2.1 and over indicate that amplification is present and to what degree. Interpretation of this data will be performed and reported from the Vysis-certified Cytogenetics laboratory at SHMC.

## References

1. Wingo, P.A., Tong, T., Bolden, S., "Cancer Statistics", 1995;45:1:8-31.
2. "Cancer Rates and Risks", 4<sup>th</sup> ed., National Institutes of Health, National Cancer Institute, 1996, p. 120.
3. Slamon, D.J., Clark, G.M., Song, S.G., Levin, W.J., Ullrich, A., McGuire, W.L. "Human breast Cancer: Correlation of relapse and survival with amplification of the her-2/neu oncogene". Science, 235:177-182, 1987.
4. Xing, W.R., Gilchrist, K.W., Harris, C.P., Samson, W., Meisner, L.F. "FISH detection of HER-2/neu oncogene amplification in early onset breast cancer". Breast Cancer Res. And Treatment 39(2):203-212, 1996.
5. Press, M.F., Bernstein, L., Thomas, P.A., Meisner, L.F., Zhou, J.Y., Ma, Y., Hung, G., Robinson, R.A., Harris, C., El-Naggar, A., Slamon, D.J., Phillips, R.N., Ross, J.S., Wolman, S.R., Flom, K.J., "Her-2/neu gene amplification characterized by fluorescence in situ hybridization: poor prognosis in node-negative breast carcinomas", J. Clinical Oncology 15(8):2894-2904, 1997.

*Provided for the clients of*

PATHOLOGY ASSOCIATES MEDICAL LABORATORIES  
PACLAB NETWORK LABORATORIES  
TRI-CITIES LABORATORY  
TREASURE VALLEY LABORATORY

*For more information, please contact  
your local representative.*

# Aneuploidy and cancer

Subrata Sen, PhD

Numeric aberrations in chromosomes, referred to as aneuploidy, is commonly observed in human cancer. Whether aneuploidy is a cause or consequence of cancer has long been debated. Three lines of evidence now make a compelling case for aneuploidy being a discrete chromosome mutation event that contributes to malignant transformation and progression process. First, precise assay of chromosome aneuploidy in several primary tumors with *in situ* hybridization and comparative genomic hybridization techniques have revealed that specific chromosome aneusomies correlate with distinct tumor phenotypes. Second, aneuploid tumor cell lines and *in vitro* transformed rodent cells have been reported to display an elevated rate of chromosome instability, thereby indicating that aneuploidy is a dynamic chromosome mutation event associated with transformation of cells. Third, and most important, a number of mitotic genes regulating chromosome segregation have been found mutated in human cancer cells, implicating such mutations in induction of aneuploidy in tumors. Some of these gene mutations, possibly allowing unequal segregations of chromosomes, also cause tumorigenic transformation of cells *in vitro*. In this review, the recent publications investigating aneuploidy in human cancers, rate of chromosome instability in aneuploidy tumor cells, and genes implicated in regulating chromosome segregation found mutated in cancer cells are discussed. *Curr Opin Oncol* 2000, 12:82-88 © 2000 Lippincott Williams & Wilkins, Inc.

The University of Texas, M.D. Anderson Cancer Center, Department of Laboratory Medicine, Houston, Texas, USA

Correspondence to Subrata Sen, PhD, The University of Texas, MD Anderson Cancer Center, Department of Laboratory Medicine, Box 054, 1615 Holcombe Blvd., Houston, TX 77030, USA; tel: 713-792-2580; fax: 713-792-4094; e-mail: ssen@mdanderson.org

*Current Opinion in Oncology* 2000 12:82-88

## Abbreviations

CGH comparative genomic hybridization  
CHE Chinese hamster embryo cells  
FISH fluorescence *in situ* hybridization  
HPRC hereditary papillary renal carcinoma  
ISH *in situ* hybridization

ISSN 1040-8746 © 2000 Lippincott Williams & Wilkins, Inc.

Cancer research over the past decade has firmly established that malignant cells accumulate a large number of genetic mutations that affect differentiation, proliferation, and cell death processes. In addition, it is also recognized that most cancers are clonal, although they display extensive heterogeneity with respect to karyotypes and phenotypes of individual clonal populations. It is estimated that numeric chromosomal imbalance, referred to as *aneuploidy*, is the most prevalent genetic change recorded among over 20,000 solid tumors analyzed thus far [1]. Phenotypic diversity of the clonal populations in individual tumors involve differences in morphology, proliferative properties, antigen expression, drug sensitivity, and metastatic potentials. It has been proposed that an underlying acquired genetic instability is responsible for the multiple mutations detected in cancer cells that lead to tumor heterogeneity and progression [2]. In a somewhat contradictory argument, it has also been suggested that clonal expansion due to selection of cells undergoing normal rates of mutation can explain malignant transformation and progression process in humans [3]. Acquired genetic instability, nonetheless, is considered important for more rapid progression of the disease [4••]. Although the original hypothesis on genetic instability in cancer primarily focused on chromosome imbalances in the form of aneuploidy in tumor cells, the actual relevance of such mutations in cancer remains a controversial issue.

Whether or not aneuploidy contributes to the malignant transformation and progression process has long been debated. A prevalent idea on genetics of cancer referred to as "somatic gene mutation hypothesis" contends that gene mutations at the nucleotide level alone can cause cancer by either activating cellular proto-oncogenes to dominant cancer causing oncogenes and/or by inactivating growth inhibitory tumor suppressor genes. In this scheme of things chromosomal instability in the form of aneuploidy is a mere consequence rather than a cause of malignant transformation and progression process.

In this review, some of the recent observations on the subject are discussed and compelling evidence is provided to suggest that aneuploidy is a distinct form of genetic instability in cancer that frequently correlates with specific phenotypes and stages of the disease. Furthermore, discrete genetic targets affecting chromosomal stability in cancer cells, recently identified, are also discussed. These data provide a new direction toward elucidating the molecular mechanisms responsi-

ble for induction of aneuploidy in cancer and may eventually be exploited as novel therapeutic targets in the future.

### Genetic alterations in cancer

Alterations in many genetic loci regulating growth, senescence, and apoptosis, identified in tumor cells, have led to the current understanding of cancer as a genetic disease. The genetic changes identified in tumors include: subtle mutations in genes at the nucleotide level; chromosomal translocations leading to structural rearrangements in genes; and numeric changes in either partial segments of chromosomes or whole chromosomes (aneuploidy) causing imbalance in gene dosage.

For the purpose of this review, both segmental and whole chromosome imbalances leading to altered DNA dosage in cancer cells are included as examples of aneuploidy.

### Incidence of aneuploidy in cancer

Evidence of aneuploidy involving one or more chromosomes have been commonly reported in human tumors. Although these observations were initially made using classic cytogenetic techniques late in a tumor's evolution and were difficult to correlate with cancer progression, more recent studies have reported association of specific nonrandom chromosome aneuploidy with different biologic properties such as loss of hormone dependence and metastatic potential [5].

Classic cytogenetic studies performed on tumor cells had serious limitations in scope because they were applicable only to those cases in which mitotic chromosomes could be obtained. Because of low spontaneous rates of cell division in primary tumors, analyses depended on cells either derived selectively from advanced metastases or those grown *in vitro* for variable periods of time. In both instances, metaphases analyzed represented only a subset of primary tumor cell population. Two major advances in cytogenetic analytic techniques, *in situ* hybridization (ISH) and comparative genomic hybridization (CGH), have allowed better resolution of chromosomal aberrations in freshly isolated tumor cells [6]. ISH analyses with chromosome-specific DNA probes, a powerful adjunct to metaphasic analysis, allows assessment of chromosomal anomalies within tumor cell populations in the contexts of whole nuclear architecture and tissue organization. CGH allows genome wide screening of chromosomal anomalies without the use of specific probes even in the absence of prior knowledge of chromosomes involved. Although both techniques have certain limitations in terms of their resolution power, they nonetheless provide a better approximation of chromosomal changes occurring among tumors of various histology, grade, and stage

compared with what was possible with the classic cytogenetic techniques. Genomic ploidy measurements have also been performed at the DNA level with flow cytometry and cytofluorometric methods. Although these assays underestimate chromosome ploidy due to a chromosomal gain occasionally masking a chromosomal loss in the same cell, several studies using these methods have supported the conclusion that DNA aneuploidy closely associates with poor prognosis in various cancers [7,8]. This discussion of some recent examples published on aneuploidy in cancer includes discussion of studies dealing with DNA ploidy measurements as well. Most of these observations are correlative without direct proof of specific involvement of genes on the respective chromosomes. Identification of putative oncogenes and tumor suppressor genes on gained and lost chromosomes in aneuploid tumors, however, are providing strong evidence that chromosomes involved in aneuploidy play a critical role in the tumorigenic process.

In renal tumors, either segmental or whole chromosome aneuploidy appears to be uniquely associated with specific histologic subtypes [9]. Tumors from patients with hereditary papillary renal carcinomas (HPRC) commonly show trisomy of chromosome 7, when analyzed by CGH. Germline mutations of a putative oncogene *MET* have been detected in patients with HPRC. A recent study [10] has demonstrated that an extra copy of chromosome 7 results in nonrandom duplication of the mutant *MET* allele in HPRC, thereby implicating this trisomy in tumorigenesis. The study suggested that mutation of *MET* may render the cells more susceptible to errors in chromosome replication, and that clonal expansion of cells harboring duplicated chromosome 7 reflects their proliferative advantage. In addition to chromosome 7, trisomy of chromosome 17 in papillary tumors and also of chromosome 8 in mesoblastic nephroma are commonly seen. Association of specific chromosome imbalances with benign and malignant forms of papillary renal tumors, therefore, not only contribute to an understanding of tumor origins and evolution, but also implicate aneuploidy of the respective chromosomes in the tumorigenic transformation process.

In colorectal tumors, chromosome aneuploidy is a common occurrence. In fact, molecular allelotyping studies have suggested that limited karyotyping data available from these tumors actually underestimate the true extent of these changes. Losses of heterozygosity reflecting loss of the maternal or paternal allele in tumors are widespread and often accompanied by a gain of the opposite allele. Therefore, for example, a tumor could lose a maternal chromosome while duplicating the same paternal chromosome, leaving the tumor cell

with a normal karyotype and ploidy but an aberrant allelotype. It has been estimated that cancer of the colon, breast, pancreas, or prostate may lose an average of 25% of its alleles. It is not unusual to discover that a tumor has lost over half of its alleles [4]. In clinical settings, DNA ploidy measurements have revealed that DNA aneuploidy indicates high risk of developing severe premalignant changes in patients with ulcerative colitis, who are known to have an increased risk of developing colorectal cancer [11]. DNA aneuploidy has been found to be one of the useful indicators of lymph node metastasis in patients with gastric carcinoma and associated with poor outcome compared with diploid cases [12,13]. CGH analyses of chromosome aneuploidy, on the other hand, was reported to correlate gain of chromosome 20q with high tumor S phase fractions and loss of 4q with low tumor apoptotic indices [14]. Aneuploidy of chromosome 4 in metastatic colorectal cancer has recently been confirmed in studies that used unbiased DNA fingerprinting with arbitrarily primed polymerase chain reactions to detect moderate gains and losses of specific chromosomal DNA sequences [15]. The molecular karyotype (amplotype) generated from colorectal cancer revealed that moderate gains of sequences from chromosomes 8 and 13 occurred in most tumors, suggesting that overrepresentation of these chromosomal regions is a critical step for metastatic colorectal cancer.

In addition to being implicated in tumorigenesis and correlated with distinct tumor phenotypes, chromosome aneuploidy has been used as a marker of risk assessment and prognosis in several other cancers. The potential value of aneuploidy as a noninvasive tool to identify individuals at high risk of developing head and neck cancer appears especially promising. Interphase fluorescence *in situ* hybridization (FISH) revealed extensive aneuploidy in tumors from patients with head and neck squamous cell carcinomas (HNSCC) and also in clinically normal distant oral regions from the same individuals [16,17]. It has been proposed that a panel of chromosome probes in FISH analyses may serve as an important tool to detect subclinical tumorigenesis and for diagnosis of residual disease. The presence of aneuploid or tetraploid populations is seen in 90% to 95% of esophageal adenocarcinomas, and when seen in conjunction with Barrett's esophagus, a premalignant condition, predicts progression of disease [18,19]. Chromosome ploidy analyses in conjunction with loss of heterozygosity and gene mutation studies in Barrett's esophagus reflect evolution of neoplastic cell lineages *in vivo* [20]. Evolution of neoplastic progeny from Barrett's esophagus following somatic genetic mutations frequently involves bifurcations and loss of heterozygosity at several chromosomal loci leading to aneuploidy and cancer. Accordingly, it is hypothesized that during

tumor cell evolution diploid cell progenitors with somatic genetic abnormalities undergo expansion with acquired genetic instability. Such instability, often manifested in the form of increased incidence of aneuploidy, enters a phase of clonal evolution beginning in premalignant cells that proceeds over a period of time and occasionally leads to malignant transformation. The clonal evolution continues even after the emergence of cancer.

The significance of DNA and chromosome aneuploidy in other human cancers continue to be evaluated. Among papillary thyroid carcinomas, aneuploid DNA content in tumor cells was reported to correlate with distant metastases, reflecting worsened prognosis [21]. Genome wide screening of follicular thyroid tumors by CGH, on the other hand, revealed frequent loss of chromosome 22 in widely invasive follicular carcinomas [22]. Chromosome copy number gains in invasive neoplasm compared with foci of ductal carcinoma *in situ* (DCIS) with similar histology have been proposed to indicate involvement of aneuploidy in progression of human breast cancer [23]. ISH analyses of cervical intraepithelial neoplasia has provided suggestive evidence that chromosomes 1, 7 and X aneusomy is associated with progression toward cervical carcinoma [24].

Although the prognostic value of numeric aberrations remains a matter of debate in human hematopoietic neoplasia, there have been recent studies to suggest that the presence of monosomy 7 defines a distinct subgroup of acute myeloid leukemia patients [25]. It is interesting in this context that therapy-related myelodysplastic syndromes have been reported to display monosomy 5 and 7 karyotypes, reflecting poor prognosis [26].

The clinical observations, mentioned previously, are supported by *in vitro* studies in human and rodent cells in which aneuploidy is induced at early stages of transformation [27,28]. It is even suggested that aneuploidy may cause cell immortalization, in some instances, that is a critical step preceding transformation.

Finally, in an interesting study to develop transgenic mouse models of human chromosomal diseases, chromosome segment specific duplication and deletions of the genome were reported to be constructed in mouse embryonic stem cells [29]. Three duplications for a portion of mouse chromosome 11 syntenic with human chromosome 17 were established in the mouse germline. Mice with 1Mb duplication developed corneal hyperplasia and thymic tumors. The findings represent the first transgenic mouse model of aneuploidy of a defined chromosome segment that documents the direct role of chromosome aneusomy in tumorigenesis.

### **Aneuploidy as "dynamic cancer-causing mutation" instead of a "consequential state" in cancer**

According to the hypothesis previously discussed, aneuploidy represents either a "gain of function" or "loss of function" mutation at the chromosome level with a causative influence on the tumorigenesis process. The hypothesis, however, is based only on circumstantial evidence even though existence of aneuploidy is correlated with different tumor phenotypes. The existence of numeric chromosomal alterations in a tumor does not mean that the change arose as a dynamic mutation due to genomic instability, because several factors could lead to consequential aneuploidy in tumors, also. Although aneuploidy as a dynamic mutation due to genomic instability in tumor cells would occur at a certain measurable rate per cell generation, a consequential state of aneuploidy in tumors may not occur at a predictable rate under similar conditions or in tumors with similar phenotypes. In addition to genomic instability, differences in environmental factors with selective pressure, could explain high incidence of aneuploidy and other somatic mutations in tumors compared with normal cells [4]. These include humoral, cell substratum, and cell-cell interaction differences between tumor and normal cell environments. It could be argued that despite similar rates of spontaneous aneuploidy induction in normal and tumor cells, the latter are selected to proliferate due to altered selective pressure in the tumor cell environment, whereas the normal cells are eliminated through activation of apoptosis. Alternatively, of course, one could postulate that selective expression or overexpression of anti-apoptotic proteins or inactivation of proapoptotic proteins in tumor cells may counteract default induction of apoptosis in G2/M phase cells undergoing missegregation of chromosomes. Recent demonstration of overexpression of a G2/M phase anti-apoptotic protein survivin in cancer cells [30] suggests that this protein may favor aberrant progression of aneuploid transformed cells through mitosis. This would then lead to proliferation of aneuploid cell lineages, which may undergo clonal evolution.

To ascertain that aneuploidy is a dynamic mutational event, various human tumor cell lines and transformed rodent cell lines have been analyzed for the rate of aneuploidy induction. When grown under controlled *in vitro* conditions, such conditions ensure that environmental factors do not influence selective proliferation of cells with chromosome instability. In one study, Lengauer *et al.* [31•] provided unequivocal evidence by FISH analyses that losses or gains of multiple chromosomes occurred in excess of  $10^{-2}$  per chromosome per generation in aneuploid colorectal cancer cell lines. The study further concluded that such chromosomal instability appeared to be a dominant trait. Using another *in*

*vitro* model system of Chinese hamster embryo (CHE) cells, Duesberg *et al.* [32•] have also obtained similar results. With clonal cultures of CHE cells, transformed with nongenotoxic chemicals and a mitotic inhibitor, these authors demonstrated that the overwhelming majority of the transformed colonies contained more than 50% aneuploid cells, indicating that aneuploidy would have originated from the same cells that underwent transformation. All the transformed colonies tested were tumorigenic. It was further documented that the ploidy factor representing the quotient of the modal chromosome number divided by the normal diploid number, in each clone, correlated directly with the degree of chromosomal instability. Therefore, chromosomal instability was found proportional to the degree of aneuploidy in the transformed cells and the authors hypothesized that aneuploidy is a unique mechanism of simultaneously altering and destabilizing, in a massive manner, the normal cellular phenotypes. In the absence of any evidence that the transforming chemicals used in the study did not induce other somatic mutations, it is difficult to rule out the contribution of such mutations in the transformation process. These results nonetheless make a strong case for aneuploidy being a dynamic chromosome mutation event intimately associated with cancer.

### **Aneuploidy versus somatic gene mutation in cancer**

The idea that numeric chromosome imbalance or aneuploidy is a direct cause of cancer was proposed at the turn of the century by Theodore Boveri [33]. However, the hypothesis was largely ignored over the last several decades in favor of the somatic gene mutation hypothesis, mentioned earlier. Evidence accumulating in the literature lately on specific chromosome aneusomies recognized in primary tumors, incidence of aneuploidy in cells undergoing transformation, and aneuploid tumor cells showing a high rate of chromosome instability have led to the rejuvenation of Boveri's hypothesis. The concept has recently been discussed as a "vintage wine in a new bottle" [34•]. The author points out that except for rare cancers caused by dominant retroviral oncogenes, diploidy does not seem to occur in solid tumors, whereas aneuploidy is a rule rather than exception in cancer.

Aneuploidy as an effective mutagenic mechanism driving tumor progression, on the other hand, is being recognized as a viable solution to the paradox that with known mutation rate in non-germline cells ( $\sim 10^{-7}$  per gene per cell generation) tumor cell lineages cannot accumulate enough mutant genes during a human lifetime [35]. The concept is gaining significant credibility since genes that potentially affect chromosome segregation were found mutated in human cancer. Some of

these genes have also been shown to have transforming capability in *in vitro* assays. Selected recent publications describing the findings are being discussed below in reference to the mitotic targets potentially involved in inducing chromosome segregation anomalies in cells.

### Potential mitotic targets and molecular mechanisms of aneuploidy

Because aneuploidy represents numeric imbalance in chromosomes, it is reasonable to expect that aneuploidy arises due to missegregation of chromosomes during cell division. There are many potential mitotic targets, which could cause unequal segregation of chromosomes (Fig. 1). Recent investigations have identified several genes involved in regulating these mitotic targets and mitotic checkpoint functions, which can be implicated in induction of aneuploidy in tumor cells. This discussion is restricted to those mitotic targets and checkpoint genes whose abnormal functioning has been observed in cancer or has been shown to cause tumorigenic transformation of cells, in recent years. The role of telomeres is discussed elsewhere in this issue. For a more detailed description of the components of mitotic machinery and their possible involvement in causing chromosome segregation abnormalities in tumor cells, readers may refer to a recently published review [36•].

Among the mitotic targets implicated in cancer, centrosome defects have been observed in a wide variety of malignant human tumors. Centrosomes play a central role in organizing the microtubule network in interphase cells and mitotic spindle during cell division. Multipolar mitotic spindles have been observed in human cancers *in situ* and abnormalities in the form of supernumerary

centrosomes, centrosomes of aberrant size and shape as well as aberrant phosphorylation of centrosome proteins have been reported in prostate, colon, brain, and breast tumors [37,38]. In view of the findings that abnormal centrosomes retain the ability to nucleate microtubules *in vitro*, it is conceivable that cells with abnormal centrosomes may missegregate chromosomes producing aneuploid cells. The molecular and genetic bases of abnormal centrosome generation and the precise pathway through which they regulate the chromosome segregation process remain to be elucidated. Recent discovery of a centrosome-associated kinase STK15/BTAK/aurora2, naturally amplified and overexpressed in human cancers, has raised the interesting possibility that aberrant expression of this kinase is critically involved in abnormal centrosome function and unequal chromosome segregation in tumor cells [39,40]. Exogenous expression of the kinase in rodent and human cells was found to correlate with an abnormal number of centrosomes, unequal partitioning of chromosomes during division, and tumorigenic transformation of cells. It is relevant in this context to mention that the *Xenopus* homologue of human STK15/BTAK/aurora2 kinase has recently been shown to phosphorylate a microtubule motor protein XHE5, the human orthologue of which is known to participate in the centrosome separation during mitosis [41]. Findings on STK15/aurora2 kinase, thus, provide an interesting lead to a possible molecular mechanism of centrosome's role in oncogenesis. Centrosomes have, of late, been implicated in oncogenesis from studies revealing supernumerary centrosomes in p53-deficient fibroblasts and overexpression of another centrosome kinase PLK1 being detected in human non-small cell lung cancer [42].

One of the critical events that ensures equal partitioning of the chromosomes during mitosis is the proper and timely separation of sister chromatids that are attached to each other and to the mitotic spindle. Untimely separation of sister chromatids has been suspected as a cause of aneuploidy in human tumors. Cohesion between sister chromatids is established during replication of chromosomes and is retained until the next metaphase/anaphase transition. It has been shown that during metaphase-anaphase transition, the anaphase promoting complex/cyclosome triggers the degradation of a group of proteins called securins that inhibit sister chromatid separation. A vertebrate securin (v-securin) has recently been identified that inhibits sister chromatid separation and is involved in transformation and tumorigenesis. Subsequent analysis revealed that the human securin is identical to the product of the gene called pituitary tumor transforming gene, which is overexpressed in some tumors and exhibits transforming activity in NIH3T3 cells. It is proposed that elevated expression of the v-securin may contribute to generation of malignant tumors due to

Figure 1. Potential mitotic targets causing aneuploidy in oncogenesis

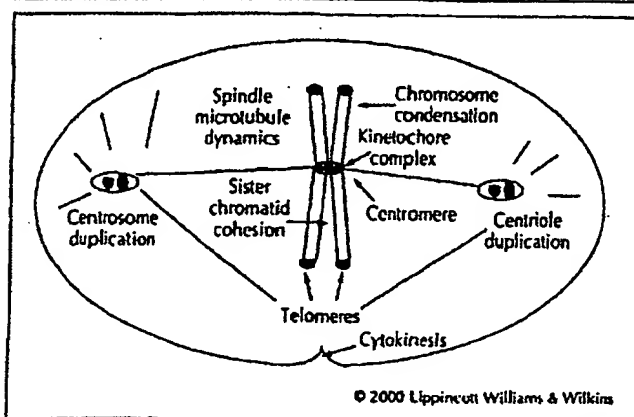


Diagram illustrates that defects in several processes involving chromosomal, spindle microtubule, and centrosomal targets, in addition to abnormal cytokinesis, may cause unequal partitioning of chromosomes during mitosis, leading to aneuploidy. Recently obtained evidence in favor of some of these possibilities is discussed in the text.

chromosome gain or loss produced by errors in chromatid separation [43•].

Normal progression through mitosis during prophase to anaphase transition is monitored at least at two checkpoints. One checkpoint operates during early prophase at G2 to metaphase progression while the second ensures proper segregation of chromosomes during metaphase to anaphase transition. Several mitotic checkpoint genes responding to mitotic spindle defects have been identified in yeast. The metaphase-anaphase transition is delayed following activation of this checkpoint during which kinetochores remain unattached to the spindle. The signal is transmitted through a kinetochore protein complex consisting of Mps1p and several Mad and Bub proteins [44]. It is expected that for unequal chromosome segregation to be perpetuated through cell proliferation cycles giving rise to aneuploidy, checkpoint controls have to be abrogated.

Following this logic, Vogelstein *et al.* [45•] hypothesized that aneuploid tumors would reveal mutation in mitotic spindle checkpoint genes. Subsequent studies by these investigators have proven the validity of this hypothesis and a small fraction of human colorectal cancers have revealed the presence of mutations in either hBub1 or hBubR1 checkpoint genes. It was further revealed that mutant BUB1 could function in a dominant negative manner conferring an abnormal spindle checkpoint when expressed exogenously. Inactivation of spindle checkpoint function in virally induced leukemia has also recently been documented following the finding that hMAD1 checkpoint protein is targeted by the Tax protein of the human T-cell leukemia virus type 1. Abrogation of hMAD1 function leads to multinucleation and aneuploidy [46].

In addition to mitotic spindle checkpoint defects, failed DNA damage checkpoint function in yeast is frequently associated with aberrant chromosome segregation as well. It, therefore, appears intriguing yet relevant that the human *BRCA1* gene, proposed to be involved in DNA damage checkpoint function, when mutated by a targeted deletion of exon 11 led to defective G2/M cell cycle checkpoint function and genetic instability in mouse embryonic fibroblasts [47]. The cells revealed multiple functional centrosomes and unequal chromosome segregation and aneuploidy. Although the molecular basis for these abnormalities is not known at this time, it raises the interesting possibility that such an aneuploidy-driven mechanism may be involved in tumorigenesis in individuals carrying germline mutations of *BRCA1* gene.

## Conclusion

Growing evidence from human tumor cytogenetic investigations strongly suggest that aneuploidy is associated with the development of tumor phenotypes. Clinical findings of correlation between aneuploidy and tumorigenesis are supported by studies with *in vitro* grown transformed cell lines. Molecular genetic analyses of tumor cells provide credible evidence that mutations in genes controlling chromosome segregation during mitosis play a critical role in causing chromosome instability leading to aneuploidy in cancer. Further elucidation of molecular and physiologic bases of chromosome instability and aneuploidy induction could lead to the development of new therapeutic approaches for common forms of cancer.

## Acknowledgments

The author is thankful to Drs. Bill Brinkley and Pramila Sen for discussions and advice. Help from Ms. Donna Sprabary and Ms. Hongyi Zhou in preparation of this manuscript is gratefully acknowledged. The work in the author's laboratory was supported by grants from the NIH and The University of Texas M.D. Anderson Cancer Center.

## References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

• Of special interest

•• Of outstanding interest

- 1 Heim S, Mitelman F: Cancer cytogenetics, edn 2. New York: Wiley-Liss Inc., 1995.
- 2 Nowell PC: The clonal evolution of tumor cell populations. *Science* 1976, 194:23-28.
- 3 Tomlinson IP, Nowell MR, Bodmer WF: The mutation rate and cancer. *Proc Natl Acad Sci USA* 1986, 83:14800-14803.
- 4 Lengauer C, Kinzler KW, Vogelstein B: Genetic instabilities in human cancers. *Nature* 1998, 396:643-649.
- An excellent review on the significance and possible mechanisms of genetic instability in cancer.
- 5 Hoppner GH, Miller FR: The cellular basis of tumor progression. *Int Rev Cytol* 1998, 177:1-56.
- 6 Wolman SR: Chromosomal markers: signposts on the road to understanding neoplastic disease. *Diag Cytopath* 1998, 18:18-23.
- 7 Ross JS: DNA ploidy and cell cycle analysis in cancer diagnosis and prognosis. *Oncology* 1998, 10:887-890.
- 8 Magennis DP: Nuclear DNA in histological and cytological specimens: measurement and prognostic significance. *Br J Biomed Sci* 1997, 54:140-148.
- 9 Fletcher JA: Renal and bladder cancers. In: *Human Cytogenetic Cancer Markers*. Edited by Wolman SR, Sell S. Totowa, NJ: Humana Press; 1997:169-202.
- 10 Zhuang Z, Park WS, Pack S, Schmidt L, Vortmeyer AO, Pak E, et al.: Trisomy 7-harboring non-random duplication of the mutant MET allele in hereditary papillary renal carcinomas. *Nat Genet* 1998, 20:66-69.
- 11 Lindberg JO, Stenling RB, Rutegard JN: DNA aneuploidy as a marker of premalignancy in surveillance of patients with ulcerative colitis. *Br J Surg* 1999, 86:947-950.
- 12 Sasaki O, Kido K, Nagahama S: DNA ploidy, Ki-67 and p53 as indicators of lymph node metastasis in early gastric carcinoma. *Anal Quant Cytol Histol* 1999, 21:85-88.
- 13 Abad M, Ciudad J, Rincon MR, Silva I, Paz-Bouza JI, Lopez A, et al.: DNA aneuploidy by flow cytometry is an independent prognostic factor in gastric cancer. *Anal Cell Path* 1998, 16:223-231.
- 14 DeAngelis PM, Clausen OP, Schjorberg A, Stokke T: Chromosomal gains and losses in primary colorectal carcinomas detected by CGH and their

## 88 Cancer biology

- associations with tumour DNA ploidy, genotypes and phenotypes. *Br J Cancer* 1999, 80:526-535.
- 15 Malkhosyan S, Yasuda J, Scoto JL, Sekiya T, Yokota J, Perucho M: Molecular karyotype (amplotype) of metastatic colorectal cancer by unbiased arbitrarily primed PCR DNA fingerprinting. *Proc Natl Acad Sci (USA)* 1998, 95:10170-10175.
  - 16 Ai H, Barrera JE, Pan Z, Meyers AD, Varela-Garcia M: Identification of individuals at high risk for head and neck carcinogenesis using chromosome aneuploidy detected by fluorescence in situ hybridization. *Mut Res* 1999, 439:223-232.
  - 17 Barrera JE, Ai H, Pan Z, Meyers AD, Varela-Garcia M: Malignancy detection by molecular cytogenetics in clinically normal mucosa adjacent to head and neck tumors. *Arch Otolaryngol Head Neck Surg* 1998, 124:847-851.
  - 18 Galipeau PC, Cowan DS, Sanchez CA, Barrett MT, Emond MJ, Levine DS, et al.: 17p (p53) allelic loss, 4N (G2/tetraploid) populations, and progression to aneuploidy in Barrett's esophagus. *Proc Natl Acad Sci USA* 1998, 95:7081-7084.
  - 19 Teodori L, Gohde W, Persiani M, Ferrario F, Tirindelli Danesi D, Scarpignato C, et al.: DNA/protein flow cytometry as a predictive marker of malignancy in dysplasia-free Barrett's esophagus: thirteen-year follow up study on a cohort of patients. *Cytometry* 1998, 34:257-263.
  - 20 Barrett MT, Sanchez CA, Prevo LJ, Wong DJ, Galipeau PC, Paulson TG, et al.: Evolution of neoplastic cell lineages in Barrett esophagus. *Nat Genet* 1999, 22:106-109.
  - 21 Sturgis CD, Caraway NP, Johnston DA, Sherman SI, Kidd L, Katz RL: Image analysis of papillary thyroid carcinoma fine needle aspirates: significant association between aneuploidy and death from disease. *Cancer* 1999, 87:155-160.
  - 22 Hemmer S, Wasenius VM, Knuutila S, Joensuu H, Franssila K: Comparison of benign and malignant follicular thyroid tumours by comparative genomic hybridization. *Br J Cancer* 1998, 78:1012-1017.
  - 23 Mendelin J, Grayson M, Wallis T, Vischer DW: Analysis of chromosome aneuploidy in breast carcinoma progression by using fluorescence in situ hybridization. *Lab Inv* 1999, 79:387-393.
  - 24 Bulten J, Poddighe PJ, Robben JC, Gemmink JH, deWilde PC, Hanselaar GAGJM: Interphase cytogenetic analysis of cervical intraepithelial neoplasia. *Am J Pathol* 1998, 152:495-503.
  - 25 Krauter J, Ganser A, Bergmann L, Raghavachar A, Hoelzer D, Libbert M, et al.: Association between structural and numerical chromosomal aberrations in acute myeloblastic leukemia: a study by RT-PCR and FISH in 447 patients with de novo AML. *Ann Hematol* 1999, 78:265-269.
  - 26 Van Den Neste E, Louvraux I, Michaux JL, Delannoy A, Michaux L, Hagemeijer A, et al.: Myelodysplastic syndrome with monosomy 5 and/or 7 following therapy with 2-chloro-2'-deoxyadenosine. *Br J Hematol* 1999, 105:268-270.
  - 27 Namba M, Mihara K, Fushimi K: immortalization of human cells and its mechanisms. *Crit Rev Oncog* 1998, 7:19-31.
  - 28 Li R, Yerganian G, Duesberg P, Kraemer A, Weller A, Rausch C, Hehlmann R: Aneuploidy correlated 100% with chemical transformation of Chinese hamster cells. *Proc Natl Acad Sci USA* 1997, 94:14506-14511.
  - 29 Liu P, Zhang H, McLellan A, Vogel H, Bradley A: Embryonic lethality and tumorigenesis caused by segmental aneuploidy on mouse chromosome 11. *Genetics* 1998, 150:1155-1168.
  - 30 Li F, Ambrosini G, Chu EY, Plescia J, Tognin S, Marchisio PC, Altieri DC: Control of apoptosis and mitotic spindle checkpoint survivin. *Nature* 1998, 395:580-584.
  - 31 Lengauer C, Kinzler KW, Vogelstein B: Genetic instability in colorectal cancers. *Nature* 1997, 386:623-627. Demonstrates chromosomal instability in aneuploid colorectal tumor cells.
  - 32 Duesberg P, Rausch C, Rausnick D, Hehlmann R: Genetic instability of cancer cells is proportional to their degree of aneuploidy. *Proc Natl Acad Sci USA* 1998, 95:13692-13697. Correlates aneuploidy and transformation in *in vitro* grown CHE cells.
  - 33 Boveri T: *Zur Frage der Entstehung maligner Tumoren*. Jena, Verlag von Gustav Fischer, 1914.
  - 34 Bialy H: Aneuploidy and cancer: vintage wine in a new bottle? *Nat Biotech* 1998, 16:137-138. Discusses the significance of aneuploidy and gene mutations in cancer.
  - 35 Orr-Weaver TL, Weinberg RA: A checkpoint on the road to cancer. *Nature* 1998, 392:223-224.
  - 36 Pihan GA, Dosssey SJ: The mitotic machinery as a source of genetic instability in cancer. *Semin Cancer Biol* 1999, 9:289-302. Describes various components and regulatory mechanisms of mitotic machinery and possible mechanisms of chromosome missegregation in cancer.
  - 37 Pihan GA, Purohit A, Wallace J, Knecht H, Udda B, Queensberry P, Dosssey SJ: Centrosome defects and genetic instability in malignant tumors. *Cancer Res* 1998, 58:3974-3985.
  - 38 Lingle WL, Lutz WH, Ingle JN, Mathis NJ, Safiari JL: Centrosome hypertrophy in human breast tumors: implications for genomic stability and cell polarity. *Proc Natl Acad Sci USA* 1998, 95:2950-2955.
  - 39 Zhou H, Kuang J, Zhong L, Kuo WL, Gray JW, Sahin A, et al.: Tumor amplified kinase STK15/BTAK induces centrosome amplification, aneuploidy and transformation. *Nat Genet* 1998, 20:189-193. Describes oncogenic property of centrosome associated STK15/aurora2 kinase and its involvement in aneuploidy induction.
  - 40 Bischoff JR, Anderson L, Shu Y, Morse K, Ng J, Chen CS, et al.: A homologue of *Drosophila* aurora kinase is oncogenic and amplified in human colorectal cancers. *EMBO J* 1998, 17:3052-3065. Describes oncogenic property of STK15/aurora2 kinase and involvement in colorectal cancers.
  - 41 Giet R, Uzbekov R, Cubizolles F, Le Guellec K, Prigent C: The xenopus laevis aurora related protein kinase pEq2 associates with and phosphorylates the Kinesin related protein X1Eq5. *J Biol Chem* 1999, 274:15005-15013.
  - 42 Zimmerman W, Sparks C, Dosssey S: Amorphous no longer: the centrosome comes into focus. *Curr Opin Cell Biol* 1998, 11:122-128.
  - 43 Zou H, McGarry TJ, Bernal T, Kirschner MW: Identification of a vertebrate sister chromatid separation inhibitor involved in transformation and tumorigenesis. *Science* 1999, 285:418-421. Demonstrates transforming and tumorigenic function of a gene inhibiting sister chromatid separation.
  - 44 Hardwick KG: The spindle checkpoint. *Trends Genet* 1998, 14:1-4.
  - 45 Cahill DP, Lengauer C, Yu J, Riggins GJ, Willson JKV, et al.: Mutations of mitotic checkpoint genes in human cancers. *Nature* 1998, 392:300-303. Describes mitotic checkpoint gene mutations in human colorectal cancers showing chromosome instability.
  - 46 Jin DY, Spencer F, Jeang KT: Human T cell leukemia virus type I oncoprotein Tax targets the human mitotic checkpoint protein MAD1. *Cell* 1998, 33:81-91.
  - 47 Xu X, Weaver Z, Unke SP, Li C, Gotay J, Wang XW, et al.: Centrosome amplification and a defective G2-M cell cycle checkpoint induce genetic instability in BRCA1 exon 11 isoform deficient cells. *Mol Cell* 1999, 3:388-395.

# Genome-wide Study of Gene Copy Numbers, Transcripts, and Protein Levels in Pairs of Non-invasive and Invasive Human Transitional Cell Carcinomas\*

Torben F. Ørntoft‡§, Thomas Thykjaer¶, Frederic M. Waldman||, Hans Wolf\*\*, and Julio E. Celis‡‡

Gain and loss of chromosomal material is characteristic of bladder cancer, as well as malignant transformation in general. The consequences of these changes at both the transcription and translation levels is at present unknown partly because of technical limitations. Here we have attempted to address this question in pairs of non-invasive and invasive human bladder tumors using a combination of technology that included comparative genomic hybridization, high density oligonucleotide array-based monitoring of transcript levels (5600 genes), and high resolution two-dimensional gel electrophoresis. The results showed that there is a gene dosage effect that in some cases superimposes on other regulatory mechanisms. This effect depended ( $p < 0.015$ ) on the magnitude of the comparative genomic hybridization change. In general (18 of 23 cases), chromosomal areas with more than 2-fold gain of DNA showed a corresponding increase in mRNA transcripts. Areas with loss of DNA, on the other hand, showed either reduced or unaltered transcript levels. Because most proteins resolved by two-dimensional gels are unknown it was only possible to compare mRNA and protein alterations in relatively few cases of well focused abundant proteins. With few exceptions we found a good correlation ( $p < 0.005$ ) between transcript alterations and protein levels. The implications, as well as limitations, of the approach are discussed. *Molecular & Cellular Proteomics* 1:37–45, 2002.

Aneuploidy is a common feature of most human cancers (1), but little is known about the genome-wide effect of this

From the ‡Department of Clinical Biochemistry, Molecular Diagnostic Laboratory and \*\*Department of Urology, Aarhus University Hospital, Skejby, DK-8200 Aarhus N, Denmark, ¶AROS Applied Biotechnology ApS, Gustav Wiedsvej 10, DK-8000 Aarhus C, Denmark, ||UCSF Cancer Center and Department of Laboratory Medicine, University of California, San Francisco, CA 94143-0808, and ‡‡Institute of Medical Biochemistry and Danish Centre for Human Genome Research, Ole Worms Allé 170, Aarhus University, DK-8000 Aarhus C, Denmark

Received, September 26, 2001, and in revised form, November 7, 2001

Published, MCP Papers in Press, November 13, 2001, DOI 10.1074/mcp.M100019-MCP200

phenomenon at both the transcription and translation levels. High throughput array studies of the breast cancer cell line BT474 has suggested that there is a correlation between DNA copy numbers and gene expression in highly amplified areas (2), and studies of individual genes in solid tumors have revealed a good correlation between gene dose and mRNA or protein levels in the case of c-erb-B2, *cyclin d1*, *ems1*, and N-myc (3–5). However, a high cyclin D1 protein expression has been observed without simultaneous amplification (4), and a low level of c-myc copy number increase was observed without concomitant c-myc protein overexpression (6).

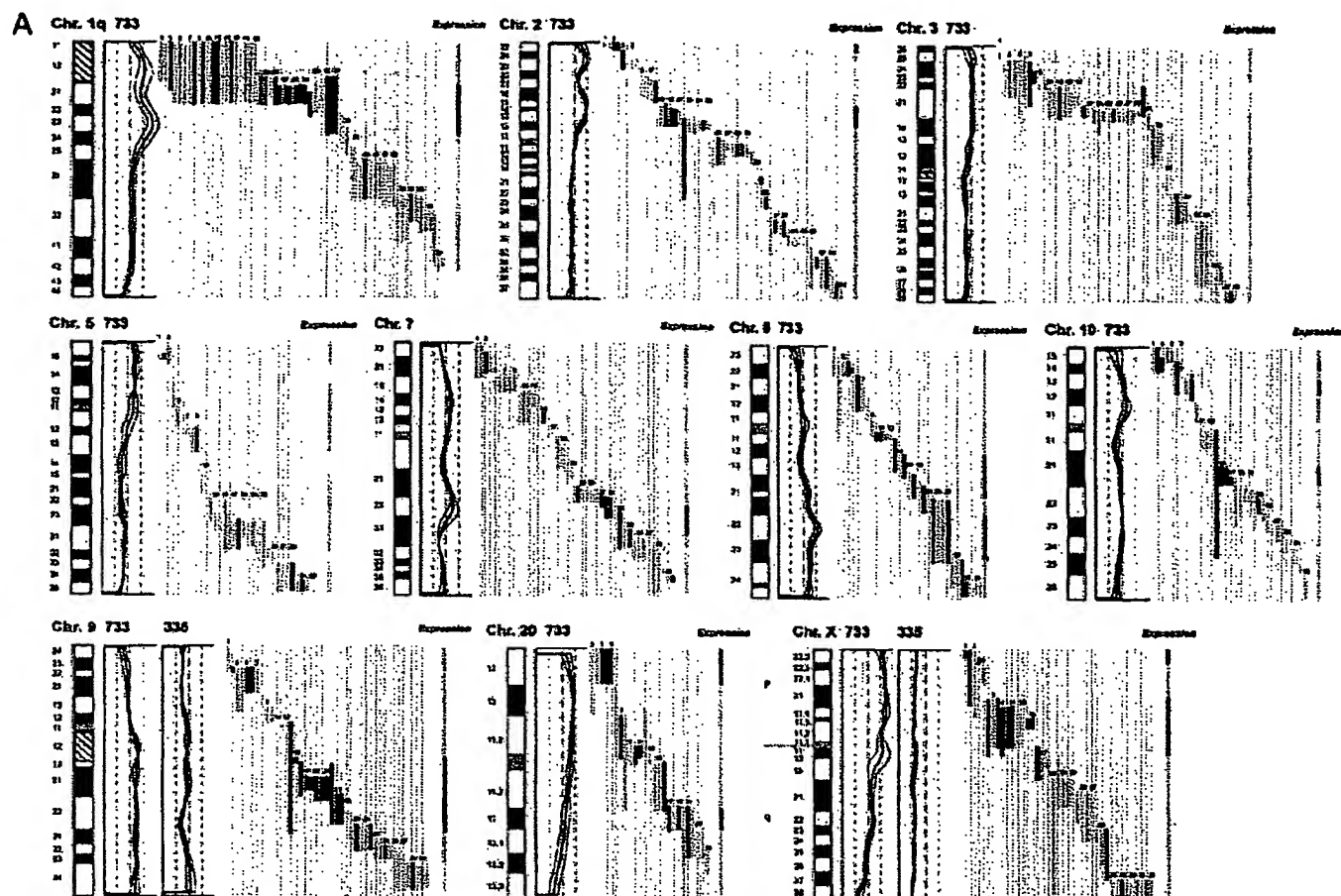
In human bladder tumors, karyotyping, fluorescent *in situ* hybridization, and comparative genomic hybridization (CGH)<sup>1</sup> have revealed chromosomal aberrations that seem to be characteristic of certain stages of disease progression. In the case of non-invasive pTa transitional cell carcinomas (TCCs), this includes loss of chromosome 9 or parts of it, as well as loss of Y in males. In minimally invasive pT1 TCCs, the following alterations have been reported: 2q–, 11p–, 1q+, 11q13+, 17q+, and 20q+ (7–12). It has been suggested that these regions harbor tumor suppressor genes and oncogenes; however, the large chromosomal areas involved often contain many genes, making meaningful predictions of the functional consequences of losses and gains very difficult.

In this investigation we have combined genome-wide technology for detecting genomic gains and losses (CGH) with gene expression profiling techniques (microarrays and proteomics) to determine the effect of gene copy number on transcript and protein levels in pairs of non-invasive and invasive human bladder TCCs.

## EXPERIMENTAL PROCEDURES

**Material**—Bladder tumor biopsies were sampled after informed consent was obtained and after removal of tissue for routine pathology examination. By light microscopy tumors 335 and 532 were staged by an experienced pathologist as pTa (superficial papillary),

<sup>1</sup> The abbreviations used are: CGH, comparative genomic hybridization; TCC, transitional cell carcinoma; LOH, loss of heterozygosity; PA-FABP, psoriasis-associated fatty acid-binding protein; 2D, two-dimensional.



**FIG. 1.** DNA copy number and mRNA expression level. Shown from left to right are chromosome (Chr.), CGH profiles, gene location and expression level of specific genes, and overall expression level along the chromosome. **A**, expression of mRNA in invasive tumor 733 as compared with the non-invasive counterpart tumor 335. **B**, expression of mRNA in invasive tumor 827 compared with the non-invasive counterpart tumor 532. The average fluorescent signal ratio between tumor DNA and normal DNA is shown along the length of the chromosome (left). The bold curve in the ratio profile represents a mean of four chromosomes and is surrounded by thin curves indicating one standard deviation. The central vertical line (broken) indicates a ratio value of 1 (no change), and the vertical lines next to it (dotted) indicate a ratio of 0.5 (left) and 2.0 (right). In chromosomes where the non-invasive tumor 335 used for comparison showed alterations in DNA content, the ratio profile of that chromosome is shown to the right of the invasive tumor profile. The colored bars represent one gene each, identified by the running numbers above the bars (the name of the gene can be seen at [www.MDL.DK/sdata.html](http://www.MDL.DK/sdata.html)). The bars indicate the purported location of the gene, and the colors indicate the expression level of the gene in the invasive tumor compared with the non-invasive counterpart; >2-fold increase (black), >2-fold decrease (blue), no significant change (orange). The bar to the far right, entitled Expression shows the resulting change in expression along the chromosome; the colors indicate that at least half of the genes were up-regulated (black), at least half of the genes down-regulated (blue), or more than half of the genes are unchanged (orange). If a gene was absent in one of the samples and present in another, it was regarded as more than a 2-fold change. A 2-fold level was chosen as this corresponded to one standard deviation in a double determination of ~1800 genes. Centromeres and heterochromatic regions were excluded from data analysis.

grade I and II, respectively, tumors 733 and 827 were staged as pT1 (invasive into submucosa), 733 was staged as solid, and 827 was staged as papillary, both grade III.

**mRNA Preparation**—Tissue biopsies, obtained fresh from surgery, were embedded immediately in a sodium-guanidinium thiocyanate solution and stored at  $-80^{\circ}\text{C}$ . Total RNA was isolated using the RNeasy B RNA isolation method (WAK-Chemie Medical GMBH). poly(A)<sup>+</sup> RNA was isolated by an oligo(dT) selection step (Oligotex mRNA kit; Qiagen).

**cRNA Preparation**—1  $\mu\text{g}$  of mRNA was used as starting material. The first and second strand cDNA synthesis was performed using the SuperScript<sup>®</sup> choice system (Invitrogen) according to the manufacturer's instructions but using an oligo(dT) primer containing a T7 RNA polymerase binding site. Labeled cRNA was prepared using the ME-GAscript<sup>®</sup> *in vitro* transcription kit (Ambion). Biotin-labeled CTP and

UTP (Enzo) was used, together with unlabeled NTPs in the reaction. Following the *in vitro* transcription reaction, the unincorporated nucleotides were removed using RNeasy columns (Qiagen).

**Array Hybridization and Scanning**—Array hybridization and scanning was modified from a previous method (13). 10  $\mu\text{g}$  of cRNA was fragmented at  $94^{\circ}\text{C}$  for 35 min in buffer containing 40 mM Tris acetate, pH 8.1, 100 mM KOAc, 30 mM MgOAc. Prior to hybridization, the fragmented cRNA in a 6 $\times$  SSPE-T hybridization buffer (1 M NaCl, 10 mM Tris, pH 7.6, 0.005% Triton), was heated to  $95^{\circ}\text{C}$  for 5 min, subsequently cooled to  $40^{\circ}\text{C}$ , and loaded onto the Affymetrix probe array cartridge. The probe array was then incubated for 16 h at  $40^{\circ}\text{C}$  at constant rotation (60 rpm). The probe array was exposed to 10 washes in 6 $\times$  SSPE-T at  $25^{\circ}\text{C}$  followed by 4 washes in 0.5 $\times$  SSPE-T at  $50^{\circ}\text{C}$ . The biotinylated cRNA was stained with a streptavidin-phycoerythrin conjugate, 10  $\mu\text{g}/\text{ml}$  (Molecular Probes) in 6 $\times$  SSPE-T

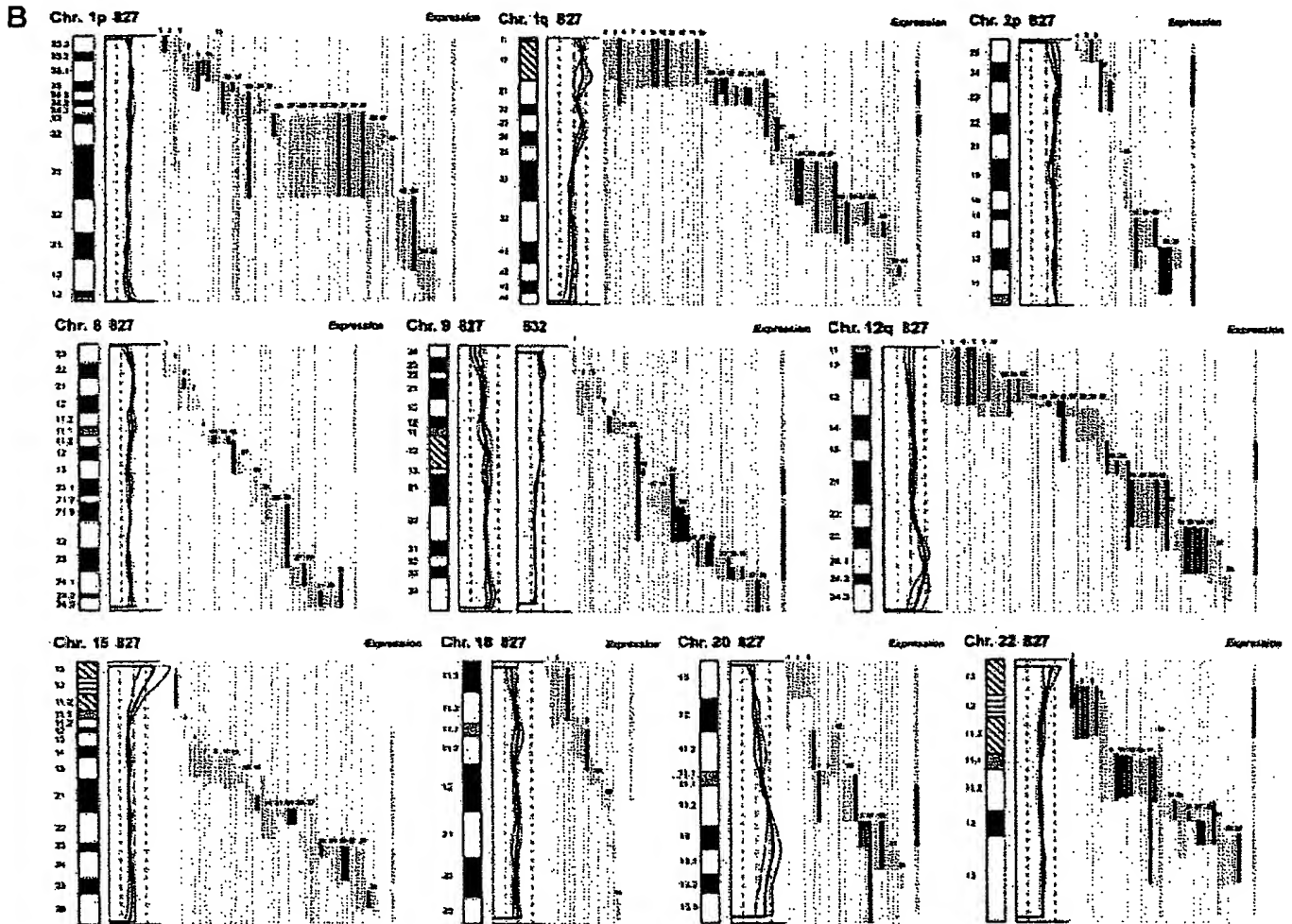


FIG. 1—continued

for 30 min at 25 °C followed by 10 washes in 6× SSPE-T at 25 °C. The probe arrays were scanned at 560 nm using a confocal laser scanning microscope (made for Affymetrix by Hewlett-Packard). The readings from the quantitative scanning were analyzed by Affymetrix gene expression analysis software.

**Microsatellite Analysis**—Microsatellite Analysis was performed as described previously (14). Microsatellites were selected by use of [www.ncbi.nlm.nih.gov/genemap98](http://www.ncbi.nlm.nih.gov/genemap98), and primer sequences were obtained from the genome data base at [www.gdb.org](http://www.gdb.org). DNA was extracted from tumor and blood and amplified by PCR in a volume of 20  $\mu$ l for 35 cycles. The amplicons were denatured and electrophoresed for 3 h in an ABI Prism 377. Data were collected in the Gene Scan program for fragment analysis. Loss of heterozygosity was defined as less than 33% of one allele detected in tumor amplicons compared with blood.

**Proteomic Analysis**—TCCs were minced into small pieces and homogenized in a small glass homogenizer in 0.5 ml of lysis solution. Samples were stored at -20 °C until use. The procedure for 2D gel electrophoresis has been described in detail elsewhere (15, 16). Gels were stained with silver nitrate and/or Coomassie Brilliant Blue. Proteins were identified by a combination of procedures that included microsequencing, mass spectrometry, two-dimensional gel Western immunoblotting, and comparison with the master two-dimensional gel image of human keratinocyte proteins; see [biobase.dk/cgi-bin/celis](http://biobase.dk/cgi-bin/celis).

**CGH**—Hybridization of differentially labeled tumor and normal DNA to normal metaphase chromosomes was performed as described previously (10). Fluorescein-labeled tumor DNA (200 ng), Texas Red-

labeled reference DNA (200 ng), and human Cot-1 DNA (20  $\mu$ g) were denatured at 37 °C for 5 min and applied to denatured normal metaphase slides. Hybridization was at 37 °C for 2 days. After washing, the slides were counterstained with 0.15  $\mu$ g/ml 4,6-diamidino-2-phenylindole in an anti-fade solution. A second hybridization was performed for all tumor samples using fluorescein-labeled reference DNA and Texas Red-labeled tumor DNA (inverse labeling) to confirm the aberrations detected during the initial hybridization. Each CGH experiment also included a normal control hybridization using fluorescein- and Texas Red-labeled normal DNA. Digital image analysis was used to identify chromosomal regions with abnormal fluorescence ratios, indicating regions of DNA gains and losses. The average green:red fluorescence intensity ratio profiles were calculated using four images of each chromosome (eight chromosomes total) with normalization of the green:red fluorescence intensity ratio for the entire metaphase and background correction. Chromosome identification was performed based on 4,6-diamidino-2-phenylindole banding patterns. Only images showing uniform high intensity fluorescence with minimal background staining were analyzed. All centromeres, p arms of acrocentric chromosomes, and heterochromatic regions were excluded from the analysis.

## RESULTS

**Comparative Genomic Hybridization**—The CGH analysis identified a number of chromosomal gains and losses in the

TABLE I  
Correlation between alterations detected by CGH and by expression monitoring

Top, CGH used as independent variable (if CGH alteration – what expression ratio was found); bottom, altered expression used as independent variable (if expression alteration – what CGH deviation was found).

CGH alterations	Tumor 733 vs. 335 Expression change clusters	Concordance	CGH alterations	Tumor 827 vs. 532 Expression change clusters	Concordance
13 Gain	10 Up-regulation 0 Down-regulation 3 No change	77%	10 Gain	8 Up-regulation 0 Down-regulation 2 No change	80%
10 Loss	1 Up-regulation 5 Down-regulation 4 No change	50%	12 Loss	3 Up-regulation 2 Down-regulation 7 No change	17%
Expression change clusters	Tumor 733 vs. 335 CGH alterations	Concordance	Expression change clusters	Tumor 827 vs. 532 CGH alterations	Concordance
16 Up-regulation	11 Gain 2 Loss 3 No change	69%	17 Up-regulation	10 Gain 5 Loss 2 No change	59%
21 Down-regulation	1 Gain 8 Loss 12 No change	38%	9 Down-regulation	0 Gain 3 Loss 6 No change	33%
15 No change	3 Gain 3 Loss 9 No change	60%	21 No change	1 Gain 3 Loss 17 No change	81%

two invasive tumors (stage pT1, TCCs 733 and 827), whereas the two non-invasive papillomas (stage pTa, TCCs 335 and 532) showed only 9p–, 9q22–q33–, and X–, and 7+, 9q–, and Y–, respectively. Both invasive tumors showed changes (1q22–24+, 2q14.1–qter–, 3q12–q13.3–, 6q12–q22–, 9q34+, 11q12–q13+, 17+, and 20q11.2–q12+) that are typical for their disease stage, as well as additional alterations, some of which are shown in Fig. 1. Areas with gains and losses deviated from the normal copy number to some extent, and the average numerical deviation from normal was 0.4-fold in the case of TCC 733 and 0.3-fold for TCC 827. The largest changes, amounting to at least a doubling of chromosomal content, were observed at 1q23 in TCC 733 (Fig. 1A) and 20q12 in TCC 827 (Fig. 1B).

**mRNA Expression in Relation to DNA Copy Number**—The mRNA levels from the two invasive tumors (TCCs 827 and 733) were compared with the two non-invasive counterparts (TCCs 532 and 335). This was done in two separate experiments in which we compared TCCs 733 to 335 and 827 to 532, respectively, using two different scaling settings for the arrays to rule out scaling as a confounding parameter. Approximately 1,800 genes that yielded a signal on the arrays were searched in the Unigene and Genemap data bases for chromosomal location, and those with a known location (1096) were plotted as bars covering their purported locus. In that way it was possible to construct a graphic presentation of DNA copy number and relative mRNA levels along the individual chromosomes (Fig. 1).

For each mRNA a ratio was calculated between the level in the invasive versus the non-invasive counterpart. Bars, which represent chromosomal location of a gene, were color-coded according to the expression ratio, and only differences larger

than 2-fold were regarded as informative (Fig. 1). The density of genes along the chromosomes varied, and areas containing only one gene were excluded from the calculations. The resolution of the CGH method is very low, and some of the outlier data may be because of the fact that the boundaries of the chromosomal aberrations are not known at high resolution.

Two sets of calculations were made from the data. For the first set we used CGH alterations as the independent variable and estimated the frequency of expression alterations in these chromosomal areas. In general, areas with a strong gain of chromosomal material contained a cluster of genes having increased mRNA expression. For example, both chromosomes 1q21–q25, 2p and 9q, showed a relative gain of more than 100% in DNA copy number that was accompanied by increased mRNA expression levels in the two tumor pairs (Fig. 1). In most cases, chromosomal gains detected by CGH were accompanied by an increased level of transcripts in both TCCs 733 (77%) and 827 (80%) (Table I, top). Chromosomal losses, on the other hand, were not accompanied by decreased expression in several cases, and were often registered as having unaltered RNA levels (Table I, top). The inability to detect RNA expression changes in these cases was not because of fewer genes mapping to the lost regions (data not shown).

In the second set of calculations we selected expression alterations above 2-fold as the independent variable and estimated the frequency of CGH alterations in these areas. As above, we found that increased transcript expression correlated with gain of chromosomal material (TCC 733, 69% and TCC 827, 59%), whereas reduced expression was often detected in areas with unaltered CGH ratios (Table I, bottom). Furthermore, as a control we looked at areas with no alter-

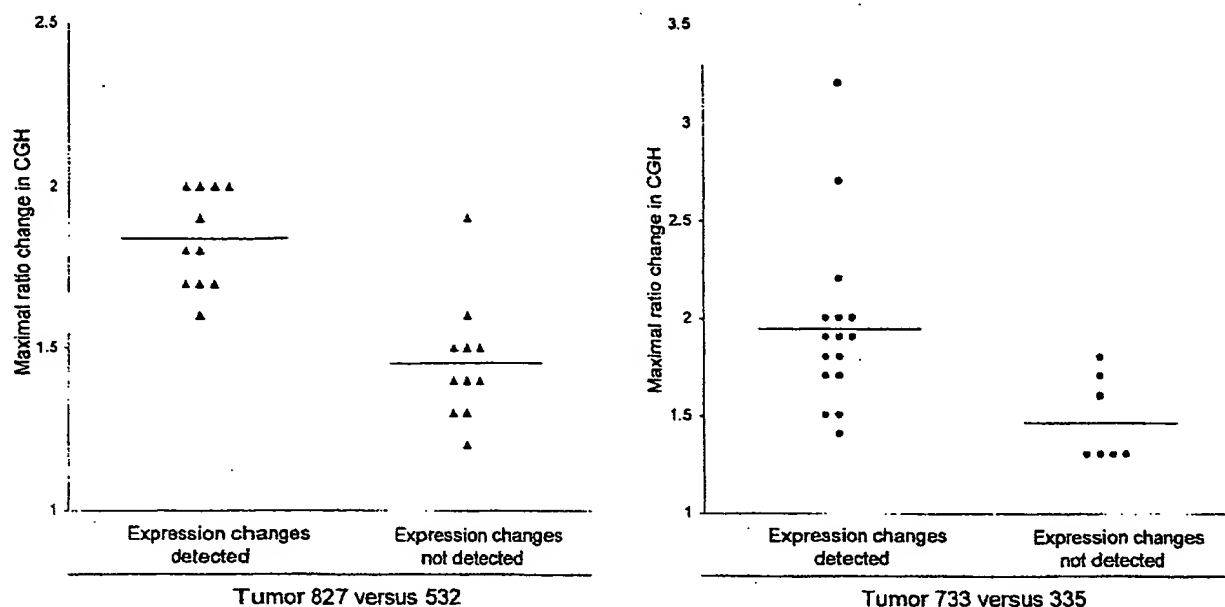


FIG. 2. Correlation between maximum CGH aberration and the ability to detect expression change by oligonucleotide array monitoring. The aberration is shown as a numerical -fold change in ratio between invasive tumors 827 (▲) and 733 (◆) and their non-invasive counterparts 532 and 335. The expression change was taken from the *Expression* line to the right in Fig. 1, which depicts the resulting expression change for a given chromosomal region. At least half of the mRNAs from a given region have to be either up- or down-regulated to be scored as an expression change. All chromosomal arms in which the CGH ratio plus or minus one standard deviation was outside the ratio value of one were included.

ation in expression. No alteration was detected by CGH in most of these areas (TCC 733, 60% and TCC 827, 81%; see Table I, bottom). Because the ability to observe reduced or increased mRNA expression clustering to a certain chromosomal area clearly reflected the extent of copy number changes, we plotted the maximum CGH aberrations in the regions showing CGH changes against the ability to detect a change in mRNA expression as monitored by the oligonucleotide arrays (Fig. 2). For both tumors TCC 733 ( $p < 0.015$ ) and TCC 827 ( $p < 0.00003$ ) a highly significant correlation was observed between the level of CGH ratio change (reflecting the DNA copy number) and alterations detected by the array based technology (Fig. 2). Similar data were obtained when areas with altered expression were used as independent variables. These areas correlated best with CGH when the CGH ratio deviated 1.6- to 2.0-fold (Table I, bottom) but mostly did not at lower CGH deviations. These data probably reflect that loss of an allele may only lead to a 50% reduction in expression level, which is at the cut-off point for detection of expression alterations. Gain of chromosomal material can occur to a much larger extent.

**Microsatellite-based Detection of Minor Areas of Losses—**In TCC 733, several chromosomal areas exhibiting DNA amplification were preceded or followed by areas with a normal CGH but reduced mRNA expression (see Fig. 1, TCC 733 chromosome 1q32, 2p21, and 7q21 and q32, 9q34, and 10q22). To determine whether these results were because of undetected loss of chromosomal material in these regions or

because of other non-structural mechanisms regulating transcription, we examined two microsatellites positioned at chromosome 1q25–32 and two at chromosome 2p22. Loss of heterozygosity (LOH) was found at both 1q25 and at 2p22 indicating that minor deleted areas were not detected with the resolution of CGH (Fig. 3). Additionally, chromosome 2p in TCC 733 showed a CGH pattern of gain/no change/gain of DNA that correlated with transcript increase/decrease/increase. Thus, for the areas showing increased expression there was a correlation with the DNA copy number alterations (Fig. 1A). As indicated above, the mRNA decrease observed in the middle of the chromosomal gain was because of LOH, implying that one of the mechanisms for mRNA down-regulation may be regions that have undergone smaller losses of chromosomal material. However, this cannot be detected with the resolution of the CGH method.

In both TCC 733 and TCC 827, the telomeric end of chromosome 11p showed a normal ratio in the CGH analysis; however, clusters of five and three genes, respectively, lost their expression. Two microsatellites (D11S1760, D11S922) positioned close to MUC2, IGF2, and cathepsin D indicated LOH as the most likely mechanism behind the loss of expression (data not shown).

A reduced expression of mRNA observed in TCC 733 at chromosomes 3q24, 11p11, 12p12.2, 12q21.1, and 16q24 and in TCC 827 at chromosome 11p15.5, 12p11, 15q11.2, and 18q12 was also examined for chromosomal losses using microsatellites positioned as close as possible to the gene loci

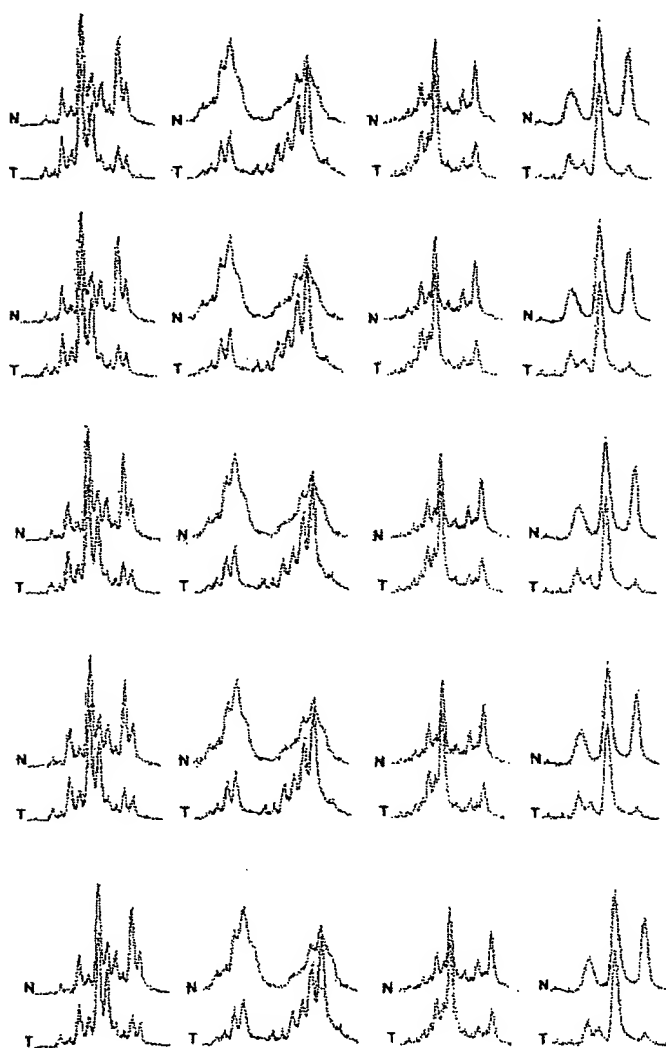


FIG. 3. Microsatellite analysis of loss of heterozygosity. Tumor 733 showing loss of heterozygosity at chromosome 1q25, detected (a) by D1S215 close to Hu class I histocompatibility antigen (gene number 38 in Fig. 1), (b) by D1S2735 close to cathepsin E (gene number 41 in Fig. 1), and (c) at chromosome 2p23 by D2S2251 close to general  $\beta$ -spectrin (gene number 11 on Fig. 1) and of (d) tumor 827 showing loss of heterozygosity at chromosome 18q12 by S18S1118 close to mitochondrial 3-oxoacyl-coenzyme A thiolase (gene number 12 in Fig. 1). The upper curves show the electropherogram obtained from normal DNA from leukocytes (N), and the lower curves show the electropherogram from tumor DNA (T). In all cases one allele is partially lost in the tumor amplicon.

showing reduced mRNA transcripts. Only the microsatellite positioned at 18q12 showed LOH (Fig. 3), suggesting that transcriptional down-regulation of genes in the other regions may be controlled by other mechanisms.

**Relation between Changes in mRNA and Protein Levels—**2D-PAGE analysis, in combination with Coomassie Brilliant Blue and/or silver staining, was carried out on all four tumors using fresh biopsy material. 40 well resolved abundant known proteins migrating in areas away from the edges of the pH

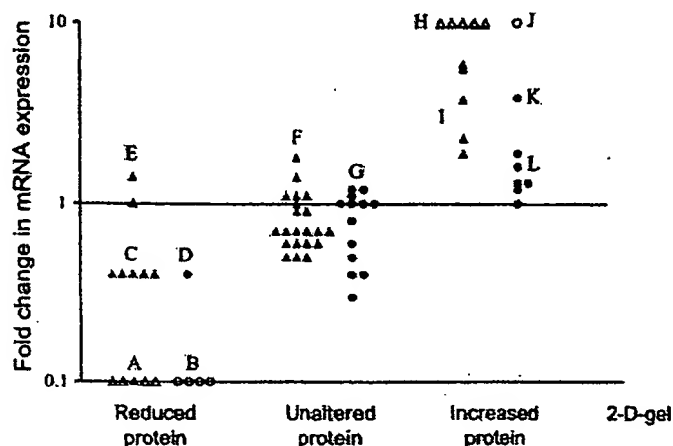
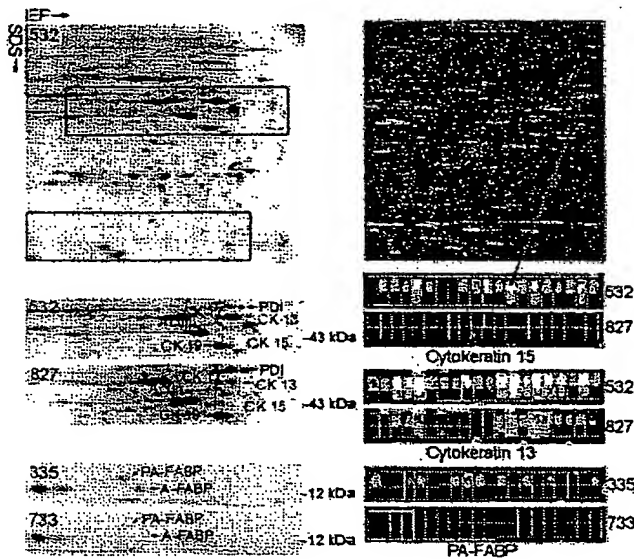


FIG. 4. Correlation between protein levels as judged by 2D-PAGE and transcript ratio. For comparison proteins were divided in three groups, unaltered in level or up- or down-regulated (horizontal axis). The mRNA ratio as determined by oligonucleotide arrays was plotted for each gene (vertical axis). ▲, mRNAs that were scored as present in both tumors used for the ratio calculation; △, mRNAs that were scored as absent in the invasive tumors (along horizontal axis) or as absent in non-invasive reference (top of figure). Two different scalings were used to exclude scaling as a confounder, TCCs 827 and 532 (▲▲) were scaled with background suppression, and TCCs 733 and 335 (●●) were scaled without suppression. Both comparisons showed highly significant ( $p < 0.005$ ) differences in mRNA ratios between the groups. Proteins shown were as follows: Group A (from left), phosphoglucosylase 1, glutathione transferase class  $\mu$  number 4, fatty acid-binding protein homologue, cytokeratin 15, and cytokeratin 13; B (from left), fatty acid-binding protein homologue, 28-kDa heat shock protein, cytokeratin 13, and calyculin; C (from left),  $\alpha$ -enolase, hnRNP B1, 28-kDa heat shock protein, 14-3-3- $\epsilon$ , and pre-mRNA splicing factor; D, mesothelial keratin K7 (type II); E (from top), glutathione S-transferase- $\pi$  and mesothelial keratin K7 (type II); F (from top and left), adenyl cyclase-associated protein, E-cadherin, keratin 19, calgizzarin, phosphoglycerate mutase, annexin IV, cytoskeletal  $\gamma$ -actin, hnRNP A1, integral membrane protein calnexin (IP90), hnRNP H, brain-type clathrin light chain-a, hnRNP F, 70-kDa heat shock protein, heterogeneous nuclear ribonucleoprotein A/B, translationally controlled tumor protein, liver glyceraldehyde-3-phosphate dehydrogenase, keratin 8, aldehyde reductase, and Na,K-ATPase  $\beta$ -1 subunit; G, (from top and left), TCP20, calgizzarin, 70-kDa heat shock protein, calnexin, hnRNP H, cytokeratin 15, ATP synthase, keratin 19, triosephosphate isomerase, hnRNP F, liver glyceraldehyde-3-phosphatase dehydrogenase, glutathione S-transferase- $\pi$ , and keratin 8; H (from left), plasma gelsolin, autoantigen calreticulin, thioredoxin, and NAD+-dependent 15 hydroxyprostaglandin dehydrogenase; I (from top), prolyl 4-hydroxylase  $\beta$ -subunit, cytokeratin 20, cytokeratin 17, prothibition, and fructose 1,6-bisphosphatase; J, annexin II; K, annexin IV; L (from top and left), 90-kDa heat shock protein, prolyl 4-hydroxylase  $\beta$ -subunit,  $\alpha$ -enolase, GRP 78, cyclophilin, and cofilin.

gradient, and having a known chromosomal location, were selected for analysis in the TCC pair 827/532. Proteins were identified by a combination of methods (see "Experimental Procedures"). In general there was a highly significant correlation ( $p < 0.005$ ) between mRNA and protein alterations (Fig. 4). Only one gene showed disagreement between transcript alteration and protein alteration. Except for a group of cyto-



**Fig. 5. Comparison of protein and transcript levels in invasive and non-invasive TCCs.** The upper part of the figure shows a 2D gel (left) and the oligonucleotide array (right) of TCC 532. The red rectangles on the upper gel highlight the areas that are compared below. Identical areas of 2D gels of TCCs 532 and 827 are shown below. Clearly, cytokeratins 13 and 15 are strongly down-regulated in TCC 827 (red annotation). The tile on the array containing probes for cytokeratin 15 is enlarged below the array (red arrow) from TCC 532 and is compared with TCC 827. The upper row of squares in each tile corresponds to perfect match probes; the lower row corresponds to mismatch probes containing a mutation (used for correction for unspecific binding). Absence of signal is depicted as black, and the higher the signal the lighter the color. A high transcript level was detected in TCC 532 (6151 units) whereas a much lower level was detected in TCC 827 (absence of signals). For cytokeratin 13, a high transcript level was also present in TCC 532 (15659 units), and a much lower level was present in TCC 827 (623 units). The 2D gels at the bottom of the figure (left) show levels of PA-FABP and adipocyte-FABP in TCCs 335 and 733 (invasive), respectively. Both proteins are down-regulated in the invasive tumor. To the right we show the array tiles for the PA-FABP transcript. A medium transcript level was detected in the case of TCC 335 (1277 units) whereas very low levels were detected in TCC 733 (166 units). IEF, isoelectric focusing.

keratins encoded by genes on chromosome 17 (Fig. 5) the analyzed proteins did not belong to a particular family. 26 well focused proteins whose genes had a known chromosomal location were detected in TCCs 733 and 335, and of these 19 correlated ( $p < 0.005$ ) with the mRNA changes detected using the arrays (Fig. 4). For example, PA-FABP was highly expressed in the non-invasive TCC 335 but lost in the invasive counterpart (TCC 733; see Fig. 5). The smaller number of proteins detected in both 733 and 335 was because of the smaller size of the biopsies that were available.

11 chromosomal regions where CGH showed aberrations that corresponded to the changes in transcript levels also showed corresponding changes in the protein level (Table II). These regions included genes that encode proteins that are found to be frequently altered in bladder cancer, namely cytokeratins 17 and 20, annexins II and IV, and the fatty acid-binding proteins PA-FABP and FBP1. Four of these proteins were encoded by genes in chromosome 17q, a frequently amplified chromosomal area in invasive bladder cancers.

#### DISCUSSION

Most human cancers have abnormal DNA content, having lost some chromosomal parts and gained others. The present study provides some evidence as to the effect of these gains and losses on gene expression in two pairs of non-invasive and invasive TCCs using high throughput expression arrays and proteomics, in combination with CGH. In general, the results showed that there is a clear individual regulation of the mRNA expression of single genes, which in some cases was superimposed by a DNA copy number effect. In most cases, genes located in chromosomal areas with gains often exhibited increased mRNA expression, whereas areas showing losses showed either no change or a reduced mRNA expression. The latter might be because of the fact that losses most often are restricted to loss of one allele, and the cut-off point for detection of expression alterations was a 2-fold change, thus being at the border of detection. In several cases, how-

**TABLE II**  
Proteins whose expression level correlates with both mRNA and gene dose changes

Protein	Chromosomal location	Tumor TCC	CGH alteration	Transcript alteration <sup>a</sup>	Protein alteration
Annexin II	1q21	733	Gain	Abs to Pres <sup>a</sup>	Increase
Annexin IV	2p13	733	Gain	3.9-Fold up	Increase
Cytokeratin 17	17q12-q21	827	Gain	3.8-Fold up	Increase
Cytokeratin 20	17q21.1	827	Gain	5.6-Fold up	Increase
(PA-)FABP	8q21.2	827	Loss	10-Fold down	Decrease
FBP1	9q22	827	Gain	2.3-Fold up	Increase
Plasma gelsolin	9q31	827	Gain	Abs to Pres	Increase
Heat shock protein 28	15q12-q13	827	Loss	2.5-Fold up	Decrease
Prohibitin	17q21	827/733	Gain	3.7-/2.5-Fold up <sup>b</sup>	Increase
Prolyl-4-hydroxyl	17q25	827/733	Gain	5.7-/1.6-Fold up	Increase
hnRNPB1	7p15	827	Loss	2.5-Fold down	Decrease

<sup>a</sup> Abs, absent; Pres, present.

<sup>b</sup> In cases where the corresponding alterations were found in both TCCs 827 and 733 these are shown as 827/733.

ever, an increase or decrease in DNA copy number was associated with *de novo* occurrence or complete loss of transcript, respectively. Some of these transcripts could not be detected in the non-invasive tumor but were present at relatively high levels in areas with DNA amplifications in the invasive tumors (e.g. in TCC 733 transcript from cellular ligand of annexin II gene (chromosome 1q21) from absent to 2670 arbitrary units; in TCC 827 transcript from small proline-rich protein 1 gene (chromosome 1q12-q21.1) from absent to 1326 arbitrary units). It may be anticipated from these data that significant clustering of genes with an increased expression to a certain chromosomal area indicates an increased likelihood of gain of chromosomal material in this area.

Considering the many possible regulatory mechanisms acting at the level of transcription, it seems striking that the gene dose effects were so clearly detectable in gained areas. One hypothetical explanation may lie in the loss of controlled methylation in tumor cells (17–19). Thus, it may be possible that in chromosomes with increased DNA copy numbers two or more alleles could be demethylated simultaneously leading to a higher transcription level, whereas in chromosomes with losses the remaining allele could be partly methylated, turning off the process (20, 21). A recent report has documented a ploidy regulation of gene expression in yeast, but in this case all the genes were present in the same ratio (22), a situation that is not analogous to that of cancer cells, which show marked chromosomal aberrations, as well as gene dosage effects.

Several CGH studies of bladder cancer have shown that some chromosomal aberrations are common at certain stages of disease progression, often occurring in more than 1 of 3 tumors. In pTa tumors, these include 9p–, 9q–, 1q+, Y– (2, 6), and in pT1 tumors, 2q–, 11p–, 11q–, 1q+, 5p+, 8q+, 17q+, and 20q+ (2–4, 6, 7). The pTa tumors studied here showed similar aberrations such as 9p– and 9q22–q33– and 9q– and Y–, respectively. Likewise, the two minimal invasive pT1 tumors showed aberrations that are commonly seen at that stage, and TCC 827 had a remarkable resemblance to the commonly seen pattern of losses and gains, such as 1q22–24 amplification (seen in both tumors), 11q14–q22 loss, the latter often linked to 17 q+ (both tumors), and 1q+ and 9p–, often linked to 20q+ and 11 q13+ (both tumors) (7–9). These observations indicate that the pairs of tumors used in this study exhibit chromosomal changes observed in many tumors, and therefore the findings could be of general importance for bladder cancer.

Considering that the mapping resolution of CGH is of about 20 megabases it is only possible to get a crude picture of chromosomal instability using this technique. Occasionally, we observed reduced transcript levels close to or inside regions with increased copy numbers. Analysis of these regions by positioning heterozygous microsatellites as close as possible to the locus showing reduced gene expression revealed loss of heterozygosity in several cases. It seems likely that multiple and different events occur along each chromosomal

arm and that the use of cDNA microarrays for analysis of DNA copy number changes will reach a resolution that can resolve these changes, as has recently been proposed (2). The outlier data were not more frequent at the boundaries of the CGH aberrations. At present we do not know the mechanism behind chromosomal aneuploidy and cannot predict whether chromosomal gains will be transcribed to a larger extent than the two native alleles. A mechanism as genetic imprinting has an impact on the expression level in normal cells and is often reduced in tumors. However, the relation between imprinting and gain of chromosomal material is not known.

We regard it as a strength of this investigation that we were able to compare invasive tumors to benign tumors rather than to normal urothelium, as the tumors studied were biologically very close and probably may represent successive steps in the progression of bladder cancer. Despite the limited amount of fresh tissue available it was possible to apply three different state of the art methods. The observed correlation between DNA copy number and mRNA expression is remarkable when one considers that different pieces of the tumor biopsies were used for the different sets of experiments. This indicates that bladder tumors are relatively homogenous, a notion recently supported by CGH and LOH data that showed a remarkable similarity even between tumors and distant metastasis (10, 23).

In the few cases analyzed, mRNA and protein levels showed a striking correspondence although in some cases we found discrepancies that may be attributed to translational regulation, post-translational processing, protein degradation, or a combination of these. Some transcripts belong to undertranslated mRNA pools, which are associated with few translationally inactive ribosomes; these pools, however, seem to be rare (24). Protein degradation, for example, may be very important in the case of polypeptides with a short half-life (e.g. signaling proteins). A poor correlation between mRNA and protein levels was found in liver cells as determined by arrays and 2D-PAGE (25), and a moderate correlation was recently reported by Ideker *et al.* (26) in yeast.

Interestingly, our study revealed a much better correlation between gained chromosomal areas and increased mRNA levels than between loss of chromosomal areas and reduced mRNA levels. In general, the level of CGH change determined the ability to detect a change in transcript. One possible explanation could be that by losing one allele the change in mRNA level is not so dramatic as compared with gain of material, which can be rather unlimited and may lead to a severalfold increase in gene copy number resulting in a much higher impact on transcript level. The latter would be much easier to detect on the expression arrays as the cut-off point was placed at a 2-fold level so as not to be biased by noise on the array. Construction of arrays with a better signal to noise ratio may in the future allow detection of lesser than 2-fold alterations in transcript levels, a feature that may facilitate the analysis of the effect of loss of chromosomal areas on transcript levels.

In eleven cases we found a significant correlation between DNA copy number, mRNA expression, and protein level. Four of these proteins were encoded by genes located at a frequently amplified area in chromosome 17q. Whether DNA copy number is one of the mechanisms behind alteration of these eleven proteins is at present unknown and will have to be proved by other methods using a larger number of samples. One factor making such studies complicated is the large extent of protein modification that occurs after translation, requiring immunoidentification and/or mass spectrometry to correctly identify the proteins in the gels.

In conclusion, the results presented in this study exemplify the large body of knowledge that may be possible to gather in the future by combining state of the art techniques that follow the pathway from DNA to protein (26). Here, we used a traditional chromosomal CGH method, but in the future high resolution CGH based on microarrays with many thousand radiation hybrid-mapped genes will increase the resolution and information derived from these types of experiments (2). Combined with expression arrays analyzing transcripts derived from genes with known locations, and 2D gel analysis to obtain information at the post-translational level, a clearer and more developed understanding of the tumor genome will be forthcoming.

**Acknowledgments**—We thank Mie Madsen, Hanne Steen, Inge Lis Thorsen, Hans Lund, Viktor Ørntoft, and Lynn Bjerke for technical help and Thomas Gingeras, Christine Harrington, and Morten Østergaard for valuable discussions.

\* This work was supported by grants from The Danish Cancer Society, the University of Aarhus, Aarhus County, Novo Nordic, the Danish Biotechnology Program, the Frenkels Foundation, the John and Birthe Meyer Foundation, and NCI, National Institutes of Health Grant CA47537. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ To whom correspondence should be addressed: Dept. of Clinical Biochemistry, Molecular Diagnostic Laboratory, Aarhus University Hospital, Skejby, DK-8200 Aarhus N, Denmark. Tel.: 45-89495100/45-86156201 (private); Fax: 45-89496018; E-mail: orntoft@kba.sks.au.dk.

## REFERENCES

- Lengauer, C., Kinzler, K. W., and Vogelstein, B. (1998) Genetic instabilities in human cancers. *Nature* **17**, 643–649
- Pollack, J. R., Perou, C. M., Alizadeh, A. A., Eisen, M. B., Pergamenschikov, A., Williams, C. F., Jeffrey, S. S., Botstein, D., and Brown, P. O. (1999) Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat. Genet.* **23**, 41–46
- de Cremoux, P., Martin, E. C., Vincent-Salomon, A., Dieras, V., Barbaroux, C., Liva, S., Pouillart, P., Sastre-Garau, X., and Magdelenat, H. (1999) Quantitative PCR analysis of c-erb B-2 (HER2/neu) gene amplification and comparison with p185(HER2/neu) protein expression in breast cancer drill biopsies. *Int. J. Cancer* **83**, 157–161
- Brugier, P. P., Tamimi, Y., Shuuring, E., and Schalken, J. (1996) Expression of cyclin D1 and EMS1 in bladder tumors; relationship with chromosome 11q13 amplifications. *Oncogene* **12**, 1747–1753
- Slavc, I., Ellenbogen, R., Jung, W. H., Vawter, G. F., Kretschmar, C., Grier, H., and Korf, B. R. (1990) *myc* gene amplification and expression in primary human neuroblastoma. *Cancer Res.* **50**, 1459–1463
- Sauter, G., Carroll, P., Moch, H., Kallioniemi, A., Kerschmann, R., Narayan, P., Mihatsch, M. J., and Waldman, F. M. (1995) *c-myc* copy number gains in bladder cancer detected by fluorescence *in situ* hybridization. *Am. J. Pathol.* **146**, 1131–1139
- Richter, J., Jiang, F., Gorog, J. P., Sartorius, G., Egenter, C., Gasser, T. C., Moch, H., Mihatsch, M. J., and Sauter, G. (1997) Marked genetic differences between stage pTa and stage pT1 papillary bladder cancer detected by comparative genomic hybridization. *Cancer Res.* **57**, 2860–2864
- Richter, J., Beffa, L., Wagner, U., Schraml, P., Gasser, T. C., Moch, H., Mihatsch, M. J., and Sauter, G. (1998) Patterns of chromosomal imbalances in advanced urinary bladder cancer detected by comparative genomic hybridization. *Am. J. Pathol.* **153**, 1615–1621
- Bruch, J., Wöhr, G., Hautmann, R., Mattfeldt, T., Bruderlein, S., Möller, P., Sauter, S., Hameister, H., Vogel, W., and Paiss, T. (1998) Chromosomal changes during progression of transitional cell carcinoma of the bladder and delineation of the amplified interval on chromosome arm 8q. *Genes Chromosomes Cancer* **23**, 167–174
- Hovey, R. M., Chu, L., Balazs, M., De Vries, S., Moore, D., Sauter, G., Carroll, P. R., and Waldman, F. M. (1998) Genetic alterations in primary bladder cancers and their metastases. *Cancer Res.* **15**, 3555–3560
- Simon, R., Burger, H., Brinkschmidt, C., Böcker, W., Hertle, L., and Terpe, H. J. (1998) Chromosomal aberrations associated with invasion in papillary superficial bladder cancer. *J. Pathol.* **185**, 345–351
- Koo, S. H., Kwon, K. C., Ihm, C. H., Jeon, Y. M., Park, J. W., and Sul, C. K. (1999) Detection of genetic alterations in bladder tumors by comparative genomic hybridization and cytogenetic analysis. *Cancer Genet. Cytogenet.* **110**, 87–93
- Wodicka, L., Dong, H., Mittmann, M., Ho, M. H., and Lockhart, D. J. (1997) Genome-wide expression monitoring in *Saccharomyces cerevisiae*. *Nat. Biotechnol.* **15**, 1359–1367
- Christensen, M., Sunde, L., Bolund, L., and Orntoft, T. F. (1999) Comparison of three methods of microsatellite detection. *Scand. J. Clin. Lab. Invest.* **59**, 167–177
- Celis, J. E., Østergaard, M., Basse, B., Celis, A., Lauridsen, J. B., Ratz, G. P., Andersen, I., Hein, B., Wolf, H., Orntoft, T. F., and Rasmussen, H. H. (1996) Loss of adipocyte-type fatty acid binding protein and other protein biomarkers is associated with progression of human bladder transitional cell carcinomas. *Cancer Res.* **56**, 4782–4790
- Celis, J. E., Ratz, G., Basse, B., Lauridsen, J. B., and Celis, A. (1994) In *Cell Biology: A Laboratory Handbook* (Celis, J. E., ed) Vol. 3, pp. 222–230, Academic Press, Orlando, FL
- Ohlsson, R., Tycko, B., and Szapienta, C. (1998) Monoallelic expression: 'there can only be one'. *Trends Genet.* **14**, 435–438
- Hollander, G. A., Zuklys, S., Morel, C., Mizoguchi, E., Mobisson, K., Simpson, S., Terhorst, C., Wishart, W., Golan, D. E., Bhan, A. K., and Burakoff, S. J. (1998) Monoallelic expression of the interleukin-2 locus. *Science* **279**, 2118–2121
- Brannan, C. I., and Bartolomei, M. S. (1999) Mechanisms of genomic imprinting. *Curr. Opin. Genet. Dev.* **9**, 164–170
- Ohlsson, R., Cui, H., He, L., Pfeifer, S., Malmikumpu, H., Jiang, S., Feinberg, A. P., and Hedborg, F. (1999) Mosaic allelic insulin-like growth factor 2 expression patterns reveal a link between Wilms' tumorigenesis and epigenetic heterogeneity. *Cancer Res.* **59**, 3889–3892
- Cui, H., Hedborg, F., He, L., Nordenskjöld, A., Sandstedt, B., Pfeifer-Ohlsson, S., and Ohlsson, R. (1997) Inactivation of H19, an imprinted and putative tumor repressor gene, is a preneoplastic event during Wilms' tumorigenesis. *Cancer Res.* **57**, 4469–4473
- Galitski, T., Saldanha, A. J., Styles, C. A., Lander, E. S., and Fink, G. R. (1999) Ploidy regulation of gene expression. *Science* **285**, 251–254
- Tsao, J., Yatabe, Y., Markl, I. D., Haijyan, K., Jones, P. A., and Shibata, D. (2000) Bladder cancer genotype stability during clinical progression. *Genes Chromosomes Cancer* **29**, 26–32
- Zong, Q., Schummer, M., Hood, L., and Morris, D. R. (1999) Messenger RNA translation state: the second dimension of high-throughput expression screening. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 10632–10636
- Anderson, L., and Seilhamer, J. (1997) Comparison of selected mRNA and protein abundances in human liver. *Electrophoresis* **18**, 533–537
- Ideker, T., Thorsson, V., Ranish, J. A., Christmas, R., Buhler, J., Eng, J. K., Bumgarner, R., Goodlett, D. R., Aebersold, R., and Hood, L. (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* **292**, 929–934

# Impact of DNA Amplification on Gene Expression Patterns in Breast Cancer<sup>1,2</sup>

Elizabeth Hyman,<sup>3</sup> Päivikki Kauraniemi,<sup>3</sup> Sampsa Hautaniemi, Maija Wolf, Spyro Mousses, Ester Rozenblum, Markus Ringnér, Guido Sauter, Outi Monni, Abdel Elkahoul, Olli-P. Kallioniemi, and Anne Kallioniemi<sup>4</sup>

Howard Hughes Medical Institute-NIH Research Scholar, Bethesda, Maryland 20892 [E. H.]; Cancer Genetics Branch, National Human Genome Research Institute, NIH, Bethesda, Maryland 20892 [E. H., P. K., S. H., M. W., S. M., E. R., M. R., A. E., O. K., A. K.]; Laboratory of Cancer Genetics, Institute of Medical Technology, University of Tampere and Tampere University Hospital, FIN-33520 Tampere, Finland [P. K., A. K.]; Signal Processing Laboratory, Tampere University of Technology, FIN-33101 Tampere, Finland [S. H.]; Institute of Pathology, University of Basel, CH-4003 Basel, Switzerland [G. S.]; and Biomedicum Biochip Center, Helsinki University Hospital, Biomedicum Helsinki, FIN-00014 Helsinki, Finland [O. M.]

## ABSTRACT

Genetic changes underlie tumor progression and may lead to cancer-specific expression of critical genes. Over 1100 publications have described the use of comparative genomic hybridization (CGH) to analyze the pattern of copy number alterations in cancer, but very few of the genes affected are known. Here, we performed high-resolution CGH analysis on cDNA microarrays in breast cancer and directly compared copy number and mRNA expression levels of 13,824 genes to quantitate the impact of genomic changes on gene expression. We identified and mapped the boundaries of 24 independent amplicons, ranging in size from 0.2 to 12 Mb. Throughout the genome, both high- and low-level copy number changes had a substantial impact on gene expression, with 44% of the highly amplified genes showing overexpression and 10.5% of the highly overexpressed genes being amplified. Statistical analysis with random permutation tests identified 270 genes whose expression levels across 14 samples were systematically attributable to gene amplification. These included most previously described amplified genes in breast cancer and many novel targets for genomic alterations, including the *HOXB7* gene, the presence of which in a novel amplicon at 17q21.3 was validated in 10.2% of primary breast cancers and associated with poor patient prognosis. In conclusion, CGH on cDNA microarrays revealed hundreds of novel genes whose overexpression is attributable to gene amplification. These genes may provide insights to the clonal evolution and progression of breast cancer and highlight promising therapeutic targets.

## INTRODUCTION

Gene expression patterns revealed by cDNA microarrays have facilitated classification of cancers into biologically distinct categories, some of which may explain the clinical behavior of the tumors (1-6). Despite this progress in diagnostic classification, the molecular mechanisms underlying gene expression patterns in cancer have remained elusive, and the utility of gene expression profiling in the identification of specific therapeutic targets remains limited.

Accumulation of genetic defects is thought to underlie the clonal evolution of cancer. Identification of the genes that mediate the effects of genetic changes may be important by highlighting transcripts that are actively involved in tumor progression. Such transcripts and their encoded proteins would be ideal targets for anticancer therapies, as demonstrated by the clinical success of new therapies against amplified oncogenes, such as *ERBB2* and *EGFR* (7, 8), in breast cancer and other solid tumors. Besides amplifications of known oncogenes, over

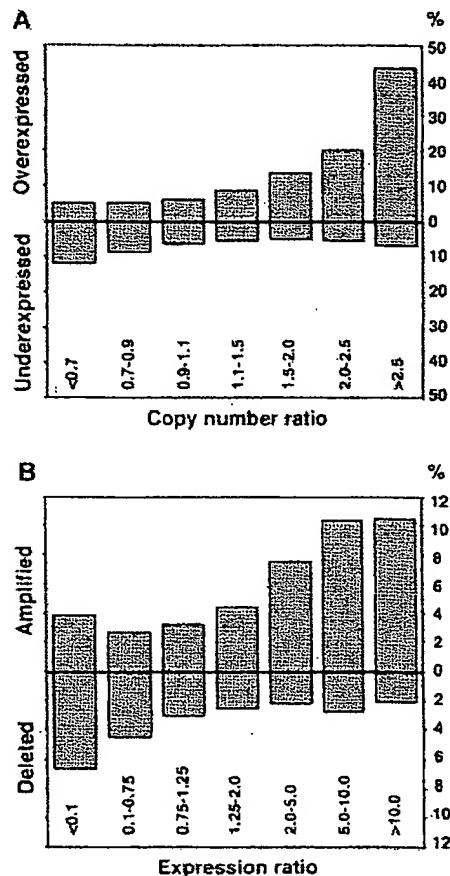


Fig. 1. Impact of gene copy number on global gene expression levels. A, percentage of over- and underexpressed genes (Y axis) according to copy number ratios (X axis). Threshold values used for over- and underexpression were  $>2.184$  (global upper 7% of the cDNA ratios) and  $<0.4826$  (global lower 7% of the expression ratios). B, percentage of amplified and deleted genes according to expression ratios. Threshold values for amplification and deletion were  $>1.5$  and  $<0.7$ .

20 recurrent regions of DNA amplification have been mapped in breast cancer by CGH<sup>5</sup> (9, 10). However, these amplicons are often large and poorly defined, and their impact on gene expression remains unknown.

We hypothesized that genome-wide identification of those gene expression changes that are attributable to underlying gene copy number alterations would highlight transcripts that are actively involved in the causation or maintenance of the malignant phenotype. To identify such transcripts, we applied a combination of cDNA and CGH microarrays to: (a) determine the global impact that gene copy number variation plays in breast cancer development and progression; and (b) identify and characterize those genes whose mRNA expres-

Received 5/29/02; accepted 8/28/02.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

<sup>1</sup> Supported in part by the Academy of Finland, Emil Aaltonen Foundation, the Finnish Cancer Society, the Pirkanmaa Cancer Society, the Pirkanmaa Cultural Foundation, the Finnish Breast Cancer Group, the Foundation for the Development of Laboratory Medicine, the Medical Research Fund of the Tampere University Hospital, the Foundation for Commercial and Technical Sciences, and the Swedish Research Council.

<sup>2</sup> Supplementary data for this article are available at Cancer Research Online (<http://cancerres.aacrjournals.org>).

<sup>3</sup> Contributed equally to this work.

<sup>4</sup> To whom requests for reprints should be addressed, at Laboratory of Cancer Genetics, Institute of Medical Technology, Lenckelilijankatu 6, FIN-33520 Tampere, Finland. Phone: 358-3247-4125; Fax: 358-3247-4168; E-mail: anne.kallioniemi@uta.fi.

<sup>5</sup> The abbreviations used are: CGH, comparative genomic hybridization; FISH, fluorescence *in situ* hybridization; RT-PCR, reverse transcription-PCR.

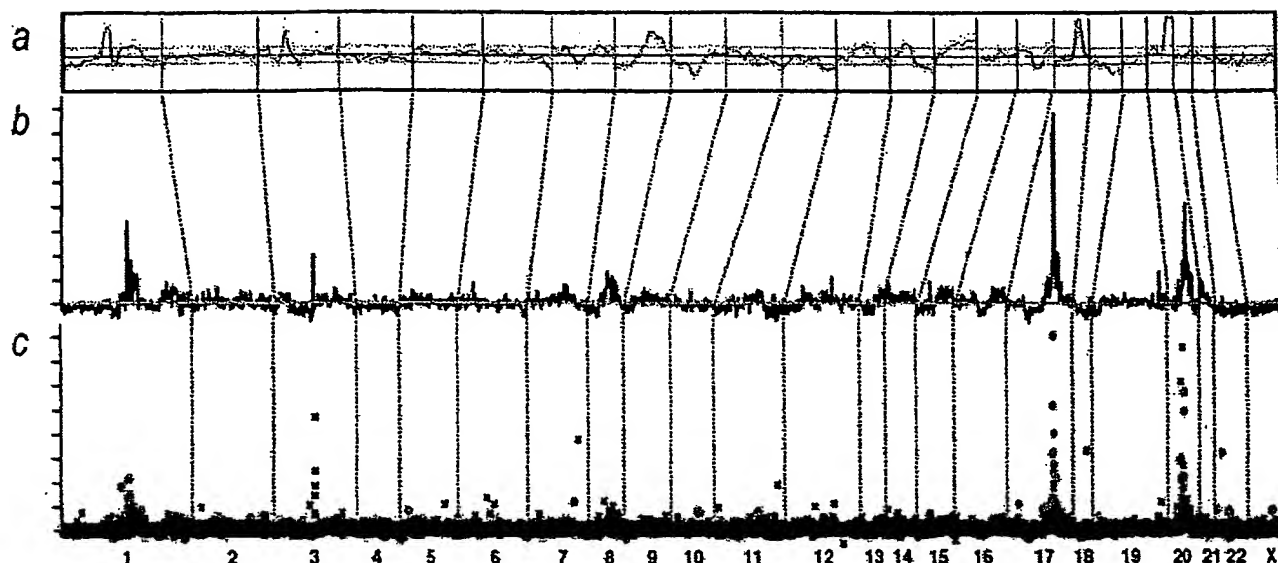


Fig. 2. Genome-wide copy number and expression analysis in the MCF-7 breast cancer cell line. *A*, chromosomal CGH analysis of MCF-7. The copy number ratio profile (blue line) across the entire genome from 1p telomere to Xq telomere is shown along with  $\pm 1$  SD (orange lines). The black horizontal line indicates a ratio of 1.0; red line, a ratio of 0.8; and green line, a ratio of 1.2. *B–C*, genome-wide copy number analysis in MCF-7 by CGH on cDNA microarray. The copy number ratios were plotted as a function of the position of the cDNA clones along the human genome. In *B*, individual data points are connected with a line, and a moving median of 10 adjacent clones is shown. Red horizontal line, the copy number ratio of 1.0. In *C*, individual data points are labeled by color coding according to cDNA expression ratios. The bright red dots indicate the upper 2%, and dark red dots, the next 5% of the expression ratios in MCF-7 cells (overexpressed genes); bright green dots indicate the lowest 2%, and dark green dots, the next 5% of the expression ratios (underexpressed genes); the rest of the observations are shown with black crosses. The chromosome numbers are shown at the bottom of the figure, and chromosome boundaries are indicated with a dashed line.

sion is most significantly associated with amplification of the corresponding genomic template.

## MATERIALS AND METHODS

**Breast Cancer Cell Lines.** Fourteen breast cancer cell lines (BT-20, BT-474, HCC1428, Hs578t, MCF7, MDA-361, MDA-436, MDA-453, MDA-468, SKBR-3, T-47D, UACC812, ZR-75-1, and ZR-75-30) were obtained from the American Type Culture Collection (Manassas, VA). Cells were grown under recommended culture conditions. Genomic DNA and mRNA were isolated using standard protocols.

**Copy Number and Expression Analyses by cDNA Microarrays.** The preparation and printing of the 13,824 cDNA clones on glass slides were performed as described (11–13). Of these clones, 244 represented uncharacterized expressed sequence tags, and the remainder corresponded to known genes. CGH experiments on cDNA microarrays were done as described (14, 15). Briefly, 20  $\mu$ g of genomic DNA from breast cancer cell lines and normal human WBCs were digested for 14–18 h with *AluI* and *RsaI* (Life Technologies, Inc., Rockville, MD) and purified by phenol/chloroform extraction. Six  $\mu$ g of digested cell line DNAs were labeled with Cy3-dUTP (Amersham Pharmacia) and normal DNA with Cy5-dUTP (Amersham Pharmacia) using the Bioprime Labeling kit (Life Technologies, Inc.). Hybridization (14, 15) and posthybridization washes (13) were done as described. For the expression analyses, a standard reference (Universal Human Reference RNA; Stratagene, La Jolla, CA) was used in all experiments. Forty  $\mu$ g of reference RNA were labeled with Cy3-dUTP and 3.5  $\mu$ g of test mRNA with Cy5-dUTP, and the labeled cDNAs were hybridized on microarrays as described (13, 15). For both microarray analyses, a laser confocal scanner (Agilent Technologies, Palo Alto, CA) was used to measure the fluorescence intensities at the target locations using the DEARRAY software (16). After background subtraction, average intensities at each clone in the test hybridization were divided by the average intensity of the corresponding clone in the control hybridization. For the copy number analysis, the ratios were normalized on the basis of the distribution of ratios of all targets on the array and for the expression analysis on the basis of 88 housekeeping genes, which were spotted four times onto the array. Low quality measurements (*i.e.*, copy number data with mean reference intensity <100 fluorescent units, and expression data with both test and reference intensity <100 fluorescent units and/or with spot size <50 units)

were excluded from the analysis and were treated as missing values. The distributions of fluorescence ratios were used to define cutpoints for increased/decreased copy number. Genes with CGH ratio >1.43 (representing the upper 5% of the CGH ratios across all experiments) were considered to be amplified, and genes with ratio <0.73 (representing the lower 5%) were considered to be deleted.

**Statistical Analysis of CGH and cDNA Microarray Data.** To evaluate the influence of copy number alterations on gene expression, we applied the following statistical approach. CGH and cDNA calibrated intensity ratios were log-transformed and normalized using median centering of the values in each cell line. Furthermore, cDNA ratios for each gene across all 14 cell lines were median centered. For each gene, the CGH data were represented by a vector that was labeled 1 for amplification (ratio, >1.43) and 0 for no amplification. Amplification was correlated with gene expression using the signal-to-noise statistics (1). We calculated a weight,  $w_g$ , for each gene as follows:

$$w_g = \frac{m_{g1} - m_{g0}}{\sigma_{g1} + \sigma_{g0}}$$

where  $m_{g1}$ ,  $\sigma_{g1}$  and  $m_{g0}$ ,  $\sigma_{g0}$  denote the means and SDs for the expression levels for amplified and nonamplified cell lines, respectively. To assess the statistical significance of each weight, we performed 10,000 random permutations of the label vector. The probability that a gene had a larger or equal weight by random permutation than the original weight was denoted by  $\alpha$ . A low  $\alpha$  (<0.05) indicates a strong association between gene expression and amplification.

**Genomic Localization of cDNA Clones and Amplicon Mapping.** Each cDNA clone on the microarray was assigned to a Unigene cluster using the Unigene Build 141.<sup>6</sup> A database of genomic sequence alignment information for mRNA sequences was created from the August 2001 freeze of the University of California Santa Cruz's GoldenPath database.<sup>7</sup> The chromosome and bp positions for each cDNA clone were then retrieved by relating these data sets. Amplicons were defined as a CGH copy number ratio >2.0 in at least two adjacent clones in two or more cell lines or a CGH ratio >2.0 in at least three adjacent clones in a single cell line. The amplicon start and end positions were

<sup>6</sup> Internet address: [http://research.nhgri.nih.gov/microarray/downloadable\\_cdna.html](http://research.nhgri.nih.gov/microarray/downloadable_cdna.html).

<sup>7</sup> Internet address: [www.genome.ucsc.edu](http://www.genome.ucsc.edu).

Table 1 Summary of independent amplicons in 14 breast cancer cell lines by CGH microarray

Location	Start (Mb)	End (Mb)	Size (Mb)
1p13	132.79	132.94	0.2
1q21	173.92	177.25	3.3
1q22	179.28	179.57	0.3
3p14	71.94	74.66	2.7
7p12.1-7p11.2	55.62	60.95	5.3
7q31	125.73	130.96	5.2
7q32	140.01	140.68	0.7
8q21.11-8q21.13	86.45	92.46	6.0
8q21.3	98.45	103.05	4.6
8q23.3-8q24.14	129.88	142.15	12.3
8q24.22	151.21	152.16	1.0
9p13	38.65	39.25	0.6
13q22-q31	77.15	81.38	4.2
16q22	86.70	87.62	0.9
17q11	29.30	30.85	1.6
17q12-q21.2	39.79	42.80	3.0
17q21.32-q21.33	52.47	55.80	3.3
17q22-q23.3	63.81	69.70	5.9
17q23.3-q24.3	69.93	74.99	5.1
19q13	40.63	41.40	0.8
20q11.22	34.59	35.85	1.3
20q13.12	44.00	45.62	1.6
20q13.12-q13.13	46.45	49.43	3.0
20q13.2-q13.32	51.32	59.12	7.8

extended to include neighboring nonamplified clones (ratio, <1.5). The amplicon size determination was partially dependent on local clone density.

**FISH.** Dual-color interphase FISH to breast cancer cell lines was done as described (17). Bacterial artificial chromosome clone RP11-361K8 was labeled with SpectrumOrange (Vysis, Downers Grove, IL), and SpectrumOrange-labeled probe for *EGFR* was obtained from Vysis. SpectrumGreen-labeled chromosome 7 and 17 centromere probes (Vysis) were used as a reference. A tissue microarray containing 612 formalin-fixed, paraffin-embedded primary breast cancers (17) was applied in FISH analyses as described (18). The use of these specimens was approved by the Ethics Committee of the University of Basel and by the NIH. Specimens containing a 2-fold or higher increase in the number of test probe signals, as compared with corresponding centromere signals, in at least 10% of the tumor cells were considered to be amplified. Survival analysis was performed using the Kaplan-Meier method and the log-rank test.

**RT-PCR.** The *HOXB7* expression level was determined relative to *GAPDH*. Reverse transcription and PCR amplification were performed using Access RT-PCR System (Promega Corp., Madison, WI) with 10 ng of mRNA as a template. *HOXB7* primers were 5'-GAGCAGAGGGACTCGGACTT-3' and 5'-GCGTCAGGTAGCGATTGTAG-3'.

## RESULTS

**Global Effect of Copy Number on Gene Expression.** 13,824 arrayed cDNA clones were applied for analysis of gene expression and gene copy number (CGH microarrays) in 14 breast cancer cell lines. The results illustrate a considerable influence of copy number on gene expression patterns. Up to 44% of the highly amplified transcripts (CGH ratio, >2.5) were overexpressed (i.e., belonged to the global upper 7% of expression ratios), compared with only 6% for genes with normal copy number levels (Fig. 1A). Conversely, 10.5% of the transcripts with high-level expression (cDNA ratio, >10) showed increased copy number (Fig. 1B). Low-level copy number increases and decreases were also associated with similar, although less dramatic, outcomes on gene expression (Fig. 1).

**Identification of Distinct Breast Cancer Amplicons.** Base-pair locations obtained for 11,994 cDNAs (86.8%) were used to plot copy number changes as a function of genomic position (Fig. 2, Supplement Fig. A). The average spacing of clones throughout the genome was 267 kb. This high-resolution mapping identified 24 independent breast cancer amplicons, spanning from 0.2 to 12 Mb of DNA (Table 1). Several amplification sites detected previously by chromosomal

CGH were validated, with 1q21, 17q12-q21.2, 17q22-q23, 20q13.1, and 20q13.2 regions being most commonly amplified. Furthermore, the boundaries of these amplicons were precisely delineated. In addition, novel amplicons were identified at 9p13 (38.65-39.25 Mb), and 17q21.3 (52.47-55.80 Mb).

**Direct Identification of Putative Amplification Target Genes.** The cDNA/CGH microarray technique enables the direct correlation of copy number and expression data on a gene-by-gene basis throughout the genome. We directly annotated high-resolution CGH plots with gene expression data using color coding. Fig. 2C shows that most of the amplified genes in the MCF-7 breast cancer cell line at 1p13, 17q22-q23, and 20q13 were highly overexpressed. A view of chromosome 7 in the MDA-468 cell line implicates *EGFR* as the most highly overexpressed and amplified gene at 7p11-p12 (Fig. 3A). In BT-474, the two known amplicons at 17q12 and 17q22-q23 contained numerous highly overexpressed genes (Fig. 3B). In addition, several genes, including the homeobox genes *HOXB2* and *HOXB7*, were highly amplified in a previously undescribed independent amplicon at 17q21.3. *HOXB7* was systematically amplified (as validated by FISH, Fig. 3B, inset) as well as overexpressed (as verified by RT-PCR, data not shown) in BT-474, UACC812, and ZR-75-30 cells. Furthermore, this novel

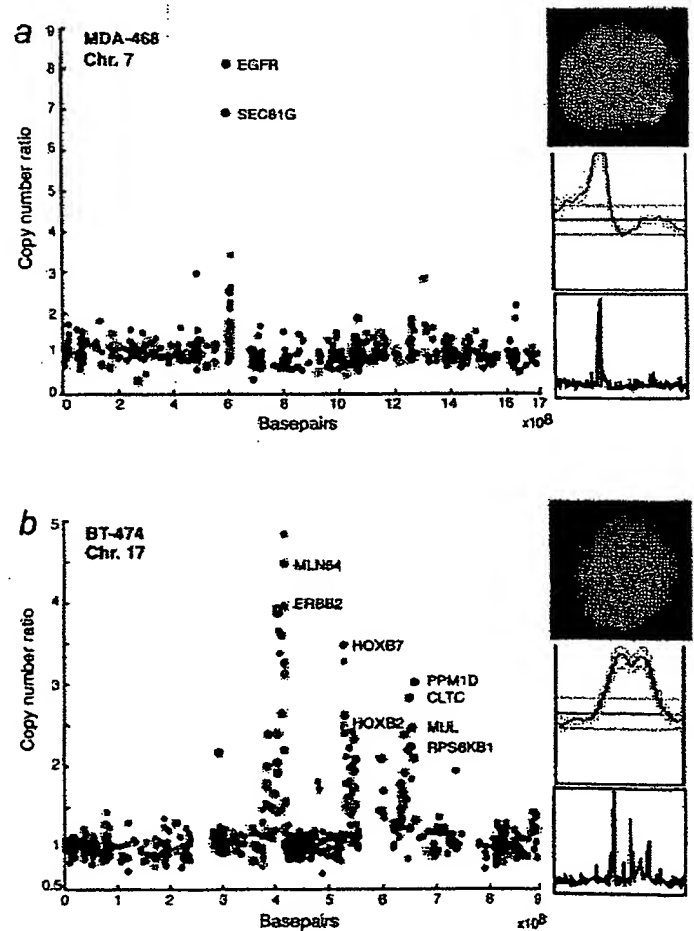
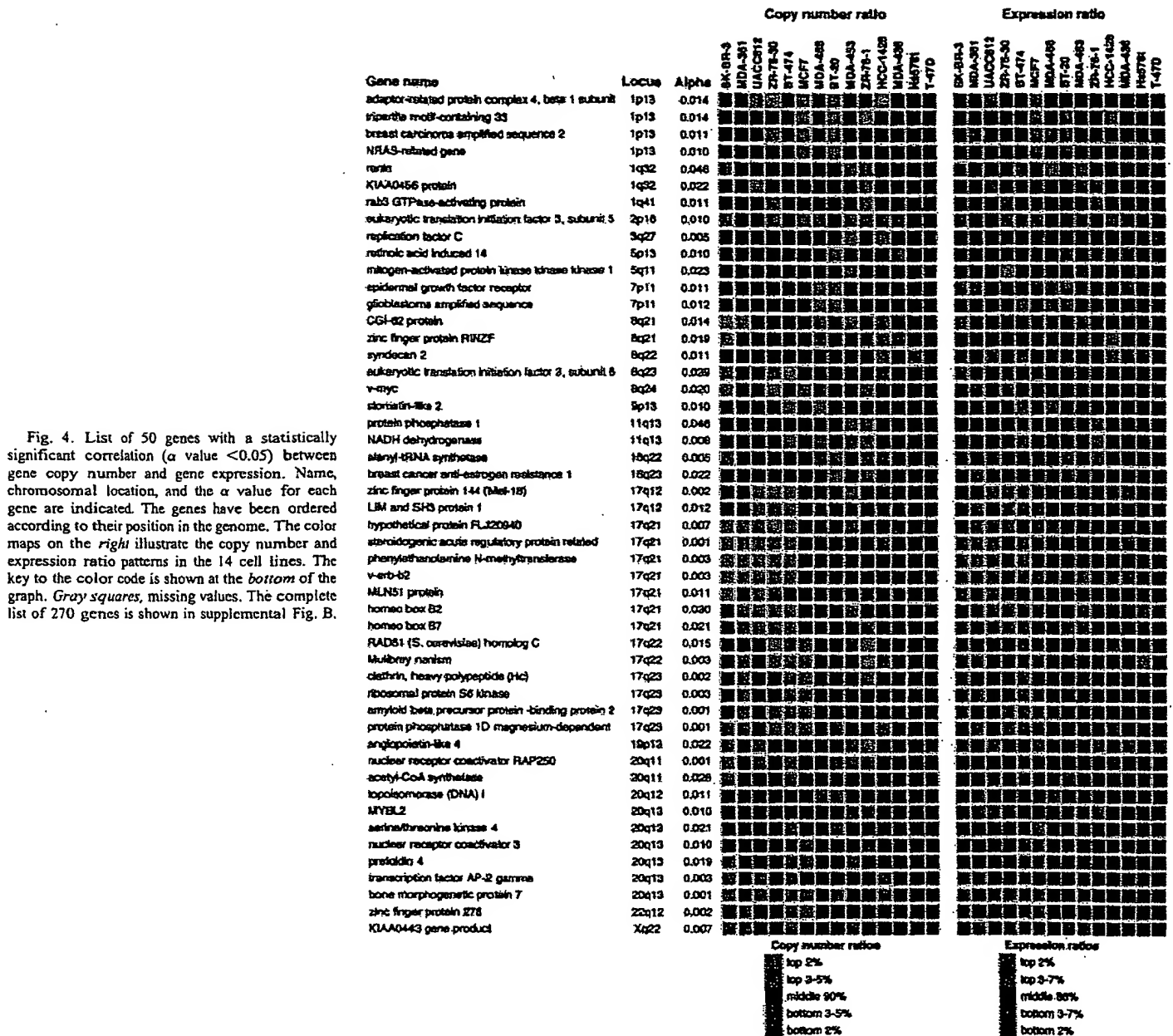


Fig. 3. Annotation of gene expression data on CGH microarray profiles. A, genes in the 7p11-p12 amplicon in the MDA-468 cell line are highly expressed (red dots) and include the *EGFR* oncogene. B, several genes in the 17q12, 17q21.3, and 17q23 amplicons in the BT-474 breast cancer cell line are highly overexpressed (red) and include the *HOXB7* gene. The data labels and color coding are as indicated for Fig. 2C. Insets show chromosomal CGH profiles for the corresponding chromosomes and validation of the increased copy number by interphase FISH using *EGFR* (red) and chromosome 7 centromere probe (green) to MDA-468 (A) and *HOXB7*-specific probe (red) and chromosome 17 centromere (green) to BT-474 cells (B).



amplification was validated to be present in 10.2% of 363 primary breast cancers by FISH to a tissue microarray and was associated with poor prognosis of the patients ( $P = 0.001$ ).

**Statistical Identification and Characterization of 270 Highly Expressed Genes in Amplicons.** Statistical comparison of expression levels of all genes as a function of gene amplification identified 270 genes whose expression was significantly influenced by copy number across all 14 cell lines (Fig. 4, Supplemental Fig. B). According to the gene ontology data,<sup>8</sup> 91 of the 270 genes represented hypothetical proteins or genes with no functional annotation, whereas 179 had associated functional information available. Of these, 151 (84%) are implicated in apoptosis, cell proliferation, signal transduction, and transcription, whereas 28 (16%) had functional annotations that could not be directly linked with cancer.

## DISCUSSION

The importance of recurrent gene and chromosome copy number changes in the development and progression of solid tumors has been characterized in >1000 publications applying CGH<sup>9</sup> (9, 10), as well as in a large number of other molecular cytogenetic, cytogenetic, and molecular genetic studies. The effects of these somatic genetic changes on gene expression levels have remained largely unknown, although a few studies have explored gene expression changes occurring in specific amplicons (15, 19–21). Here, we applied genome-wide cDNA microarrays to identify transcripts whose expression changes were attributable to underlying gene copy number alterations in breast cancer.

The overall impact of copy number on gene expression patterns was substantial with the most dramatic effects seen in the case of high-

<sup>8</sup> Internet address: <http://www.geneontology.org/>.

<sup>9</sup> Internet address: <http://www.ncbi.nlm.nih.gov/entrez>.

level copy number increase. Low-level copy number gains and losses also had a significant influence on expression levels of genes in the regions affected, but these effects were more subtle on a gene-by-gene basis than those of high-level amplifications. However, the impact of low-level gains on the dysregulation of gene expression patterns in cancer may be equally important if not more important than that of high-level amplifications. Aneuploidy and low-level gains and losses of chromosomal arms represent the most common types of genetic alterations in breast and other cancers and, therefore, have an influence on many genes. Our results in breast cancer extend the recent studies on the impact of aneuploidy on global gene expression patterns in yeast cells, acute myeloid leukemia, and a prostate cancer model system (22–24).

The CGH microarray analysis identified 24 independent breast cancer amplicons. We defined the precise boundaries for many amplicons detected previously by chromosomal CGH (9, 10, 25, 26) and also discovered novel amplicons that had not been detected previously, presumably because of their small size (only 1–2 Mb) or close proximity to other larger amplicons. One of these novel amplicons involved the homeobox gene region at 17q21.3 and led to the overexpression of the *HOXB7* and *HOXB2* genes. The homeodomain transcription factors are known to be key regulators of embryonic development and have been occasionally reported to undergo aberrant expression in cancer (27, 28). *HOXB7* transfection induced cell proliferation in melanoma, breast, and ovarian cancer cells and increased tumorigenicity and angiogenesis in breast cancer (29–32). The present results imply that gene amplification may be a prominent mechanism for overexpressing *HOXB7* in breast cancer and suggest that *HOXB7* contributes to tumor progression and confers an aggressive disease phenotype in breast cancer. This view is supported by our finding of amplification of *HOXB7* in 10% of 363 primary breast cancers, as well as an association of amplification with poor prognosis of the patients.

We carried out a systematic search to identify genes whose expression levels across all 14 cell lines were attributable to amplification status. Statistical analysis revealed 270 such genes (representing ~2% of all genes on the array), including not only previously described amplified genes, such as *HER-2*, *MYC*, *EGFR*, ribosomal protein s6 kinase, and *AIB3*, but also numerous novel genes such as *NRAS-related gene* (1p13), *syndecan-2* (8q22), and *bone morphogenic protein* (20q13.1), whose activation by amplification may similarly promote breast cancer progression. Most of the 270 genes have not been implicated previously in breast cancer development and suggest novel pathogenetic mechanisms. Although we would not expect all of them to be causally involved, it is intriguing that 84% of the genes with associated functional information were implicated in apoptosis, cell proliferation, signal transduction, transcription, or other cellular processes that could directly imply a possible role in cancer progression. Therefore, a detailed characterization of these genes may provide biological insights to breast cancer progression and might lead to the development of novel therapeutic strategies.

In summary, we demonstrate application of cDNA microarrays to the analysis of both copy number and expression levels of over 12,000 transcripts throughout the breast cancer genome, roughly once every 267 kb. This analysis provided: (a) evidence of a prominent global influence of copy number changes on gene expression levels; (b) a high-resolution map of 24 independent amplicons in breast cancer; and (c) identification of a set of 270 genes, the overexpression of which was statistically attributable to gene amplification. Characterization of a novel amplicon at 17q21.3 implicated amplification and overexpression of the *HOXB7* gene in breast cancer, including a clinical association

between *HOXB7* amplification and poor patient prognosis. Overall, our results illustrate how the identification of genes activated by gene amplification provides a powerful approach to highlight genes with an important role in cancer as well as to prioritize and validate putative targets for therapy development.

## REFERENCES

- Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P., Coller, H., Loh, M. L., Downing, J. R., Caligiuri, M. A., Bloomfield, C. D., and Lander, E. S. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* (Wash. DC), 286: 531–537, 1999.
- Alizadeh, A. A., Eisen, M. B., Davis, R. E., Ma, C., Lossos, I. S., Rosenwald, A., Boldrick, J. C., Sabet, H., Tran, T., Yu, X., et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* (Lond.), 403: 503–511, 2000.
- Bittner, M., Meltzer, P., Chen, Y., Jiang, Y., Seftor, E., Hendrix, M., Radmacher, M., Simon, R., Yakhini, Z., Ben-Dor, A., et al. Molecular classification of cutaneous malignant melanoma by gene expression profiling. *Nature* (Lond.), 406: 536–540, 2000.
- Perou, C. M., Sorlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, L. A., et al. Molecular portraits of human breast tumours. *Nature* (Lond.), 406: 747–752, 2000.
- Dhanasekaran, S. M., Barrette, T. R., Ghosh, D., Shah, R., Varambally, S., Kurachi, K., Pienta, K. J., Rubin, M. A., and Chinnaiyan, A. M. Delineation of prognostic biomarkers in prostate cancer. *Nature* (Lond.), 412: 822–826, 2001.
- Sorlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci. USA*, 98: 10869–10874, 2001.
- Ross, J. S., and Fletcher, J. A. The *HER-2/neu* oncogene: prognostic factor, predictive factor and target for therapy. *Semin. Cancer Biol.*, 9: 125–138, 1999.
- Arteaga, C. L. The epidermal growth factor receptor: from mutant oncogene in nonhuman cancers to therapeutic target in human neoplasia. *J. Clin. Oncol.*, 19: 32–40, 2001.
- Knuutila, S., Bjorkqvist, A. M., Autio, K., Tarkkanen, M., Wolf, M., Monni, O., Szymanska, J., Larramendy, M. L., Tapper, J., Pere, H., El-Rifai, W., et al. DNA copy number amplifications in human neoplasms: review of comparative genomic hybridization studies. *Am. J. Pathol.*, 152: 1107–1123, 1998.
- Knuutila, S., Autio, K., and Aalto, Y. Online access to CGH data of DNA sequence copy number changes. *Am. J. Pathol.*, 157: 689, 2000.
- DeRisi, J., Penland, L., Brown, P. O., Bittner, M. L., Meltzer, P. S., Ray, M., Chen, Y., Su, Y. A., and Trent, J. M. Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nat. Genet.*, 14: 457–460, 1996.
- Shalon, D., Smith, S. J., and Brown, P. O. A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. *Genome Res.*, 6: 639–645, 1996.
- Mousses, S., Bittner, M. L., Chen, Y., Dougherty, E. R., Baxevas, A., Meltzer, P. S., and Trent, J. M. Gene expression analysis by cDNA microarrays. In: F. J. Livesey and S. P. Hunt (eds.), *Functional Genomics*, pp. 113–137. Oxford: Oxford University Press, 2000.
- Pollack, J. R., Perou, C. M., Alizadeh, A. A., Eisen, M. B., Pergamenschikov, A., Williams, C. F., Jeffrey, S. S., Botstein, D., and Brown, P. O. Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat. Genet.*, 23: 41–46, 1999.
- Monni, O., Bärklund, M., Mousses, S., Kononen, J., Sauter, G., Heiskanen, M., Paavola, P., Avela, K., Chen, Y., Bittner, M. L., and Kallioniemi, A. Comprehensive copy number and gene expression profiling of the 17q23 amplicon in human breast cancer. *Proc. Natl. Acad. Sci. USA*, 98: 5711–5716, 2001.
- Chen, Y., Dougherty, E. R., and Bittner, M. L. Ratio-based decisions and the quantitative analysis of cDNA microarray images. *J. Biomed. Optics*, 2: 364–374, 1997.
- Bärklund, M., Forozan, F., Kononen, J., Bubendorf, L., Chen, Y., Bittner, M. L., Thorst, J., Haas, P., Bucher, C., Sauter, G., et al. Detecting activation of ribosomal protein S6 kinase by complementary DNA and tissue microarray analysis. *J. Natl. Cancer Inst.*, 92: 1252–1259, 2000.
- Andersen, C. L., Hostetter, G., Grigoryan, A., Sauter, G., and Kallioniemi, A. Improved procedure for fluorescence *in situ* hybridization on tissue microarrays. *Cytometry*, 45: 83–86, 2001.
- Kauraniemi, P., Bärklund, M., Monni, O., and Kallioniemi, A. New amplified and highly expressed genes discovered in the ERBB2 amplicon in breast cancer by cDNA microarrays. *Cancer Res.*, 61: 8235–8240, 2001.
- Clark, J., Edwards, S., John, M., Flohr, P., Gordon, T., Maillard, K., Giddings, I., Brown, C., Bagherzadeh, A., Campbell, C., Shipley, J., Wooster, R., and Cooper, C. S. Identification of amplified and expressed genes in breast cancer by comparative hybridization onto microarrays of randomly selected cDNA clones. *Genes Chromosomes Cancer*, 34: 104–114, 2002.
- Varis, A., Wolf, M., Monni, O., Vakkari, M. L., Kokkola, A., Moskaluk, C., Frierson, H., Powell, S. M., Knuutila, S., Kallioniemi, A., and El-Rifai, W. Targets of gene amplification and overexpression at 17q in gastric cancer. *Cancer Res.*, 62: 2625–2629, 2002.
- Hughes, T. R., Roberts, C. J., Dai, H., Jones, A. R., Meyer, M. R., Slade, D., Burchard, J., Dow, S., Ward, T. R., Kidd, M. J., Friend, S. H., and Marton M. J.

- Widespread aneuploidy revealed by DNA microarray expression profiling. *Nat. Genet.*, 25: 333-337, 2000.
23. Virtaneva, K., Wright, F. A., Tanner, S. M., Yuan, B., Lemon, W. J., Caligiuri, M. A., Bloomfield, C. D., de La Chapelle, A., and Krahe, R. Expression profiling reveals fundamental biological differences in acute myeloid leukemia with isolated trisomy 8 and normal cytogenetics. *Proc. Natl. Acad. Sci. USA*, 98: 1124-1129, 2001.
24. Phillips, J. L., Hayward, S. W., Wang, Y., Vasselli, J., Pavlovich, C., Padilla-Nash, H., Pezullo, J. R., Ghadimi, B. M., Grossfeld, G. D., Rivera, A., Linchan, W. M., Cunha, G. R., and Ried, T. The consequences of chromosomal aneuploidy on gene expression profiles in a cell line model for prostate carcinogenesis. *Cancer Res.*, 61: 8143-8149, 2001.
25. Bärthel, M., Tirkkonen, M., Forozan, F., Tanner, M. M., Kallioniemi, O. P., and Kallioniemi, A. Increased copy number at 17q22-q24 by CGH in breast cancer is due to high-level amplification of two separate regions. *Genes Chromosomes Cancer*, 20: 372-376, 1997.
26. Tanner, M. M., Tirkkonen, M., Kallioniemi, A., Isola, J., Kuukasjärvi, T., Collins, C., Kowbel, D., Guan, X. Y., Trent, J., Gray, J. W., Meltzer, P., and Kallioniemi O. P. Independent amplification and frequent co-amplification of three nonsyntenic regions on the long arm of chromosome 20 in human breast cancer. *Cancer Res.*, 56: 3441-3445, 1996.
27. Cillo, C., Faiella, A., Cantile, M., and Boncinelli, E. Homeobox genes and cancer. *Exp. Cell Res.*, 248: 1-9, 1999.
28. Cillo, C., Cantile, M., Faiella, A., and Boncinelli, E. Homeobox genes in normal and malignant cells. *J. Cell. Physiol.*, 188: 161-169, 2001.
29. Care, A., Silvani, A., Meccia, E., Mattia, G., Stoppacciaro, A., Parmiani, G., Peschle, C., and Colombo, M. P. HOXB7 constitutively activates basic fibroblast growth factor in melanomas. *Mol. Cell. Biol.*, 16: 4842-4851, 1996.
30. Care, A., Silvani, A., Meccia, E., Mattia, G., Peschle, C., and Colombo, M. P. Transduction of the SkBr3 breast carcinoma cell line with the HOXB7 gene induces bFGF expression, increases cell proliferation and reduces growth factor dependence. *Oncogene*, 16: 3285-3289, 1998.
31. Care, A., Felicetti, F., Meccia, E., Bottero, L., Parenza, M., Stoppacciaro, A., Peschle, C., and Colombo, M. P. HOXB7: a key factor for tumor-associated angiogenic switch. *Cancer Res.*, 61: 6532-6539, 2001.
32. Naora, H., Yang, Y. Q., Montz, F. J., Seidman, J. D., Kurman, R. J., and Roden, R. B. A serologically identified tumor antigen encoded by a homeobox gene promotes growth of ovarian epithelial cells. *Proc. Natl. Acad. Sci. USA*, 98: 4060-4065, 2001.

# Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors

Jonathan R. Pollack<sup>\*,†,‡</sup>, Therese Sørlie<sup>§</sup>, Charles M. Perou<sup>¶</sup>, Christian A. Rees<sup>||,‡</sup>, Stefanie S. Jeffrey<sup>††</sup>, Per E. Lønning<sup>††</sup>, Robert Tibshirani<sup>§§</sup>, David Botstein<sup>||</sup>, Anne-Lise Børresen-Dale<sup>§</sup>, and Patrick O. Brown<sup>††||</sup>

Departments of <sup>\*</sup>Pathology, <sup>§</sup>Genetics, <sup>††</sup>Surgery, <sup>§§</sup>Health Research and Policy, and <sup>||</sup>Biochemistry, and <sup>†</sup>Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, CA 94305; <sup>¶</sup>Department of Genetics, Norwegian Radium Hospital, Montebello, N-0310 Oslo, Norway; <sup>††</sup>Department of Medicine (Oncology), Haukeland University Hospital, N-5021 Bergen, Norway; and <sup>||</sup>Department of Genetics and Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC 27599

Contributed by Patrick O. Brown, August 6, 2002

Genomic DNA copy number alterations are key genetic events in the development and progression of human cancers. Here we report a genome-wide microarray comparative genomic hybridization (array CGH) analysis of DNA copy number variation in a series of primary human breast tumors. We have profiled DNA copy number alteration across 6,691 mapped human genes, in 44 predominantly advanced, primary breast tumors and 10 breast cancer cell lines. While the overall patterns of DNA amplification and deletion corroborate previous cytogenetic studies, the high-resolution (gene-by-gene) mapping of amplicon boundaries and the quantitative analysis of amplicon shape provide significant improvement in the localization of candidate oncogenes. Parallel microarray measurements of mRNA levels reveal the remarkable degree to which variation in gene copy number contributes to variation in gene expression in tumor cells. Specifically, we find that 62% of highly amplified genes show moderately or highly elevated expression, that DNA copy number influences gene expression across a wide range of DNA copy number alterations (deletion, low-, mid- and high-level amplification), that on average, a 2-fold change in DNA copy number is associated with a corresponding 1.5-fold change in mRNA levels, and that overall, at least 12% of all the variation in gene expression among the breast tumors is directly attributable to underlying variation in gene copy number. These findings provide evidence that widespread DNA copy number alteration can lead directly to global deregulation of gene expression, which may contribute to the development or progression of cancer.

Conventional cytogenetic techniques, including comparative genomic hybridization (CGH) (1), have led to the identification of a number of recurrent regions of DNA copy number alteration in breast cancer cell lines and tumors (2–4). While some of these regions contain known or candidate oncogenes [e.g., *FGFR1* (8p11), *MYC* (8q24), *CCND1* (11q13), *ERBB2* (17q12), and *ZNF217* (20q13)] and tumor suppressor genes [*RB1* (13q14) and *TP53* (17p13)], the relevant gene(s) within other regions (e.g., gain of 1q, 8q22, and 17q22–24, and loss of 8p) remain to be identified. A high-resolution genome-wide map, delineating the boundaries of DNA copy number alterations in tumors, should facilitate the localization and identification of oncogenes and tumor suppressor genes in breast cancer. In this study, we have created such a map, using array-based CGH (5–7) to profile DNA copy number alteration in a series of breast cancer cell lines and primary tumors.

An unresolved question is the extent to which the widespread DNA copy number changes that we and others have identified in breast tumors alter expression of genes within involved regions. Because we had measured mRNA levels in parallel in the same samples (8), using the same DNA microarrays, we had an opportunity to explore on a genomic scale the relationship between DNA copy number changes and gene expression. From

this analysis, we have identified a significant impact of widespread DNA copy number alteration on the transcriptional programs of breast tumors.

## Materials and Methods

**Tumors and Cell Lines.** Primary breast tumors were predominantly large (>3 cm), intermediate-grade, infiltrating ductal carcinomas, with more than 50% being lymph node positive. The fraction of tumor cells within specimens averaged at least 50%. Details of individual tumors have been published (8, 9), and are summarized in Table 1, which is published as supporting information on the PNAS web site, [www.pnas.org](http://www.pnas.org). Breast cancer cell lines were obtained from the American Type Culture Collection. Genomic DNA was isolated either using Qiagen genomic DNA columns, or by phenol/chloroform extraction followed by ethanol precipitation.

**DNA Labeling and Microarray Hybridizations.** Genomic DNA labeling and hybridizations were performed essentially as described in Pollack *et al.* (7), with slight modifications. Two micrograms of DNA was labeled in a total volume of 50 microliters and the volumes of all reagents were adjusted accordingly. “Test” DNA (from tumors and cell lines) was fluorescently labeled (Cy5) and hybridized to a human cDNA microarray containing 6,691 different mapped human genes (i.e., UniGene clusters). The “reference” (labeled with Cy3) for each hybridization was normal female leukocyte DNA from a single donor. The fabrication of cDNA microarrays and the labeling and hybridization of mRNA samples have been described (8).

**Data Analysis and Map Positions.** Hybridized arrays were scanned on a GenePix scanner (Axon Instruments, Foster City, CA), and fluorescence ratios (test/reference) calculated using SCANALYZE software (available at <http://rana.lbl.gov>). Fluorescence ratios were normalized for each array by setting the average log fluorescence ratio for all array elements equal to 0. Measurements with fluorescence intensities more than 20% above background were considered reliable. DNA copy number profiles that deviated significantly from background ratios measured in normal genomic DNA control hybridizations were interpreted as evidence of real DNA copy number alteration (see *Estimating Significance of Altered Fluorescence Ratios* in the supporting information). When indicated, DNA copy number profiles are displayed as a moving average (symmetric 5-nearest neighbors). Map positions for arrayed human cDNAs were assigned by

Abbreviation: CGH, comparative genomic hybridization.

<sup>†</sup>To whom reprint requests should be addressed at: Department of Pathology, Stanford University School of Medicine, CCSR Building, Room 3245A, 269 Campus Drive, Stanford, CA 94305-5176. E-mail: [pollack1@stanford.edu](mailto:pollack1@stanford.edu).

<sup>††</sup>Present address: Zymyx Inc., Hayward, CA 94545.

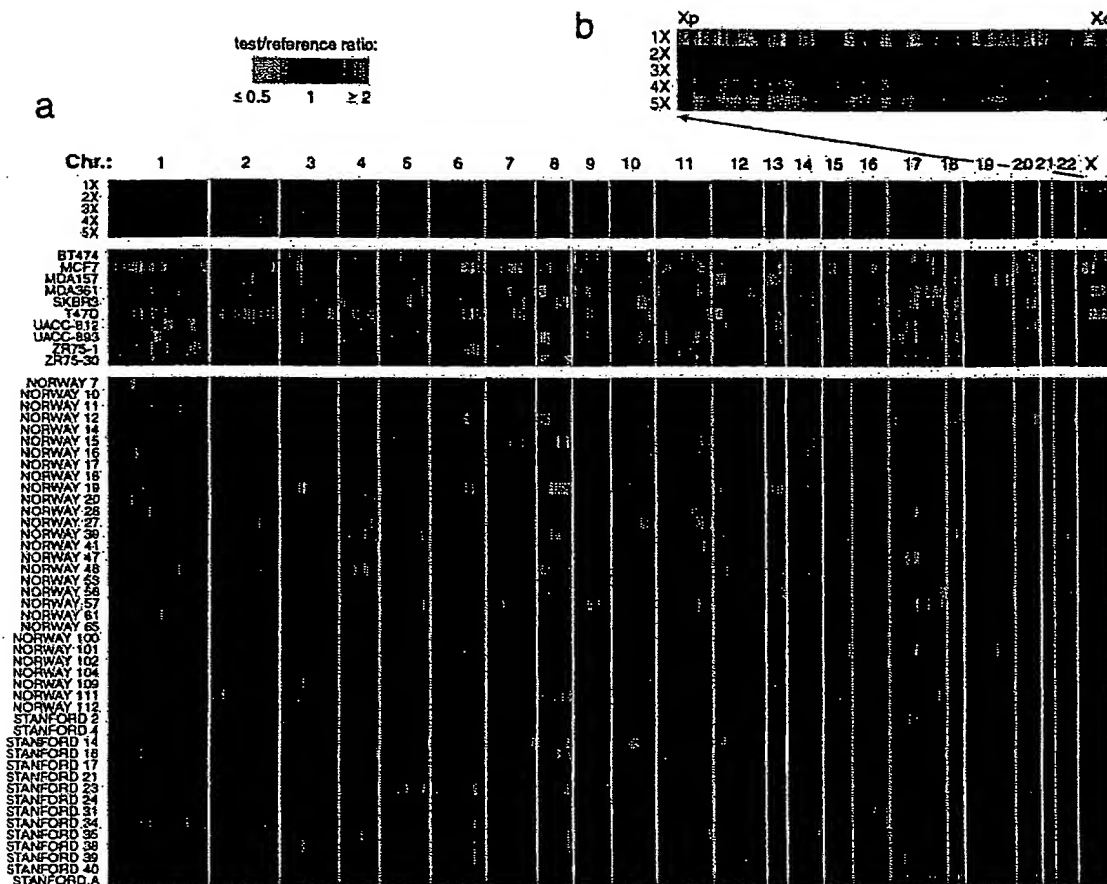


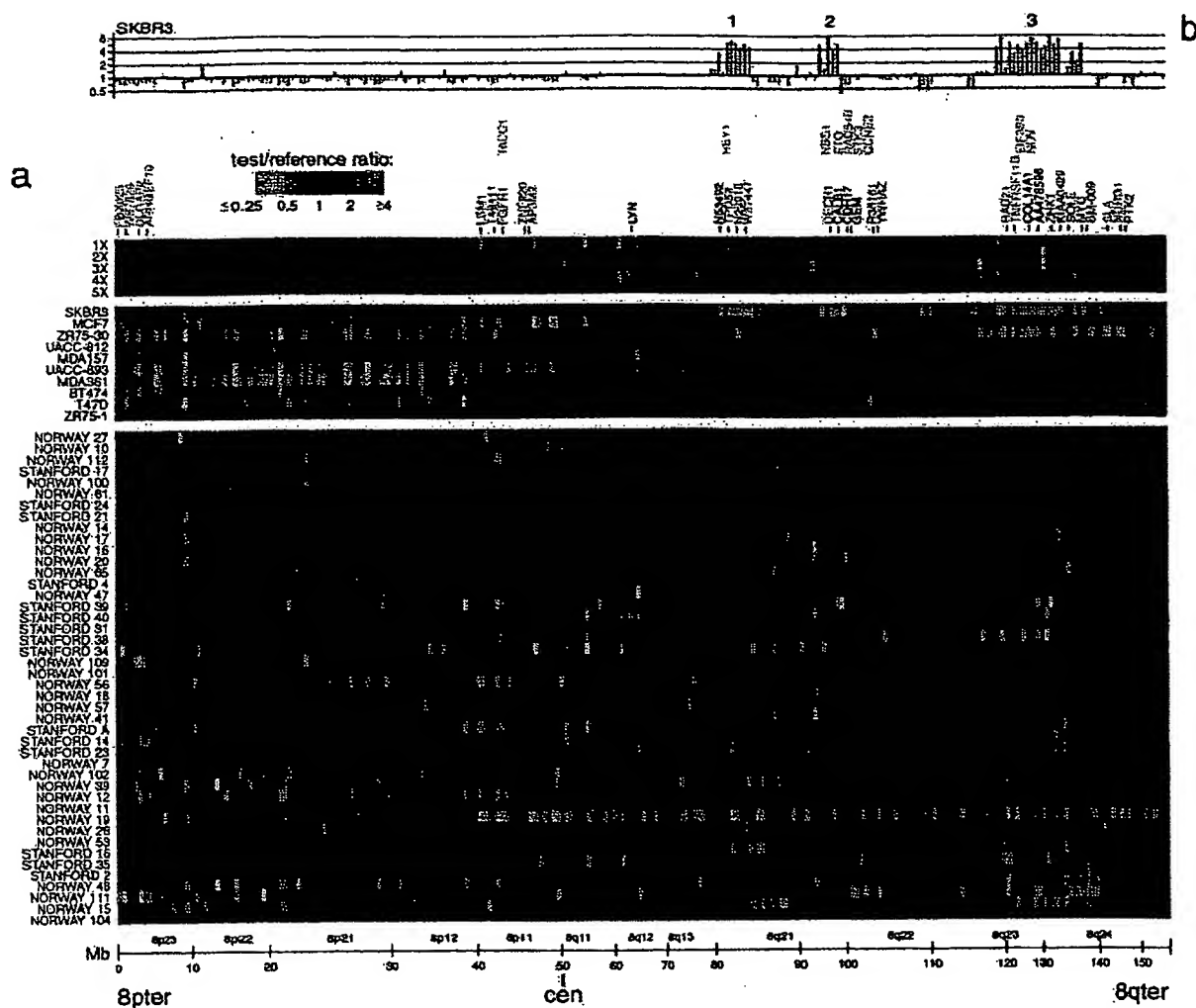
Fig. 1. Genome-wide measurement of DNA copy number alteration by array CGH. (a) DNA copy number profiles are illustrated for cell lines containing different numbers of X chromosomes, for breast cancer cell lines, and for breast tumors. Each row represents a different cell line or tumor, and each column represents one of 6,691 different mapped human genes present on the microarray, ordered by genome map position from 1pter through Xqter. Moving average (symmetric 5-nearest neighbors) fluorescence ratios (test/reference) are depicted using a log<sub>2</sub>-based pseudocolor scale (indicated), such that red luminescence reflects fold-amplification, green luminescence reflects fold-deletion, and black indicates no change (gray indicates poorly measured data). (b) Enlarged view of DNA copy number profiles across the X chromosome, shown for cell lines containing different numbers of X chromosomes.

identifying the starting position of the best and longest match of any DNA sequence represented in the corresponding UniGene cluster (10) against the "Golden Path" genome assembly (<http://genome.ucsc.edu/>; Oct 7, 2000 Freeze). For UniGene clusters represented by multiple arrayed elements, mean fluorescence ratios (for all elements representing the same UniGene cluster) are reported. For mRNA measurements, fluorescence ratios are "mean-centered" (i.e., reported relative to the mean ratio across the 44 tumor samples). The data set described here can be accessed in its entirety in the supporting information.

## Results

We performed CGH on 44 predominantly locally advanced, primary breast tumors and 10 breast cancer cell lines, using cDNA microarrays containing 6,691 different mapped human genes (Fig. 1a; also see *Materials and Methods* for details of microarray hybridizations). To take full advantage of the improved spatial resolution of array CGH, we ordered (fluorescence ratios for) the 6,691 cDNAs according to the "Golden Path" (<http://genome.ucsc.edu/>) genome assembly of the draft human genome sequences (11). In so doing, arrayed cDNAs not only themselves represent genes of potential interest (e.g., candidate oncogenes within amplicons), but also provide precise genetic landmarks for chromosomal regions of amplification and

deletion. Parallel analysis of DNA from cell lines containing different numbers of X chromosomes (Fig. 1b), as we did before (7), demonstrated the sensitivity of our method to detect single-copy loss (45, XO), and 1.5- (47,XXX), 2- (48,XXXX), or 2.5-fold (49,XXXXX) gains (also see Fig. 5, which is published as supporting information on the PNAS web site). Fluorescence ratios were linearly proportional to copy number ratios, which were slightly underestimated, in agreement with previous observations (7). Numerous DNA copy number alterations were evident in both the breast cancer cell lines and primary tumors (Fig. 1a), detected in the tumors despite the presence of euploid non-tumor cell types; the magnitudes of the observed changes were generally lower in the tumor samples. DNA copy-number alterations were found in every cancer cell line and tumor, and on every human chromosome in at least one sample. Recurrent regions of DNA copy number gain and loss were readily identifiable. For example, gains within 1q, 8q, 17q, and 20q were observed in a high proportion of breast cancer cell lines/tumors (90%/69%, 100%/47%, 100%/60%, and 90%/44%, respectively), as were losses within 1p, 3p, 8p, and 13q (80%/24%, 80%/22%, 80%/22%, and 70%/18%, respectively), consistent with published cytogenetic studies (refs. 2-4; a complete listing of gains/losses is provided in Tables 2 and 3, which are published as supporting information on the PNAS web site). The total



**Fig. 2.** DNA copy number alteration across chromosome 8 by array CGH. (a) DNA copy number profiles are illustrated for cell lines containing different numbers of X chromosomes, for breast cancer cell lines, and for breast tumors. Breast cancer cell lines and tumors are separately ordered by hierarchical clustering to highlight recurrent copy number changes. The 241 genes present on the microarrays and mapping to chromosome 8 are ordered by position along the chromosome. Fluorescence ratios (test/reference) are depicted by a log<sub>2</sub> pseudocolor scale (indicated). Selected genes are indicated with color-coded text (red, increased; green, decreased; black, no change; gray, not well measured) to reflect correspondingly altered mRNA levels (observed in the majority of the subset of samples displaying the DNA copy number change). The map positions for genes of interest that are not represented on the microarray are indicated in the row above those genes represented on the array. (b) Graphical display of DNA copy number profile for breast cancer cell line SKBR3. Fluorescence ratios (tumor/normal) are plotted on a log<sub>2</sub> scale for chromosome 8 genes, ordered along the chromosome.

number of genomic alterations (gains and losses) was found to be significantly higher in breast tumors that were high grade ( $P = 0.008$ ), consistent with published CGH data (3), estrogen receptor negative ( $P = 0.04$ ), and harboring TP53 mutations ( $P = 0.0006$ ) (see Table 4, which is published as supporting information on the PNAS web site).

The improved spatial resolution of our array CGH analysis is illustrated for chromosome 8, which displayed extensive DNA copy number alteration in our series. A detailed view of the variation in the copy number of 241 genes mapping to chromosome 8 revealed multiple regions of recurrent amplification; each of these potentially harbors a different known or previously uncharacterized oncogene (Fig. 2a). The complexity of amplicon structure is most easily appreciated in the breast cancer cell line SKBR3. Although a conventional CGH analysis of 8q in SKBR3 identified only two distinct regions of amplification (12), we observed three distinct regions of high-level amplification (labeled 1–3 in Fig. 2b). For each of these regions we can define the

boundaries of the interval recurrently amplified in the tumors we examined; in each case, known or plausible candidate oncogenes can be identified (a description of these regions, as well as the recurrently amplified regions on chromosomes 17 and 20, can be found in Figs. 6 and 7, which are published as supporting information on the PNAS web site).

For a subset of breast cancer cell lines and tumors (4 and 37, respectively), and a subset of arrayed genes (6,095), mRNA levels were quantitatively measured in parallel by using cDNA microarrays (8). The parallel assessment of mRNA levels is useful in the interpretation of DNA copy number changes. For example, the highly amplified genes that are also highly expressed are the strongest candidate oncogenes within an amplicon. Perhaps more significantly, our parallel analysis of DNA copy number changes and mRNA levels provides us the opportunity to assess the global impact of widespread DNA copy number alteration on gene expression in tumor cells.

A strong influence of DNA copy number on gene expression is evident in an examination of the pseudocolor representations

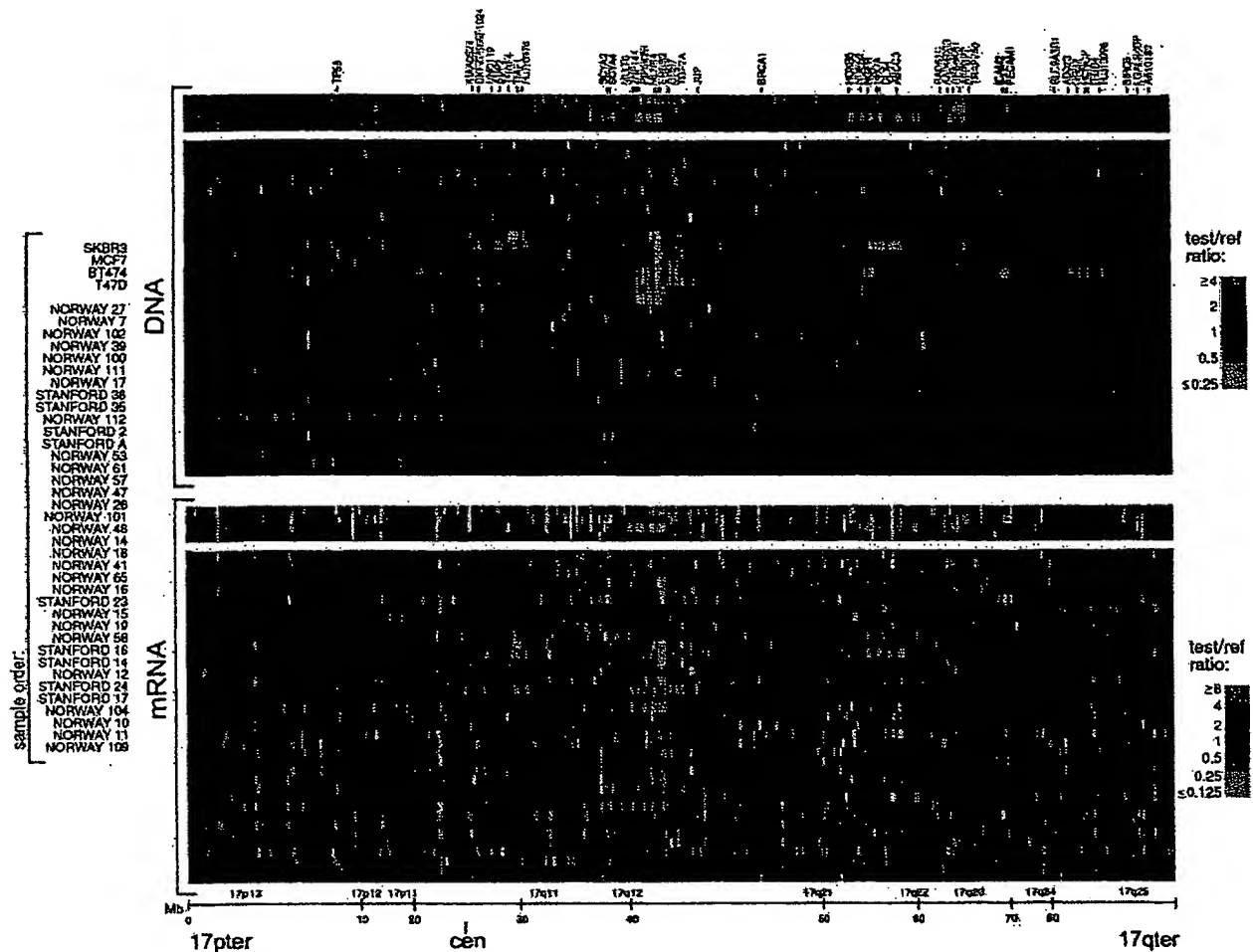
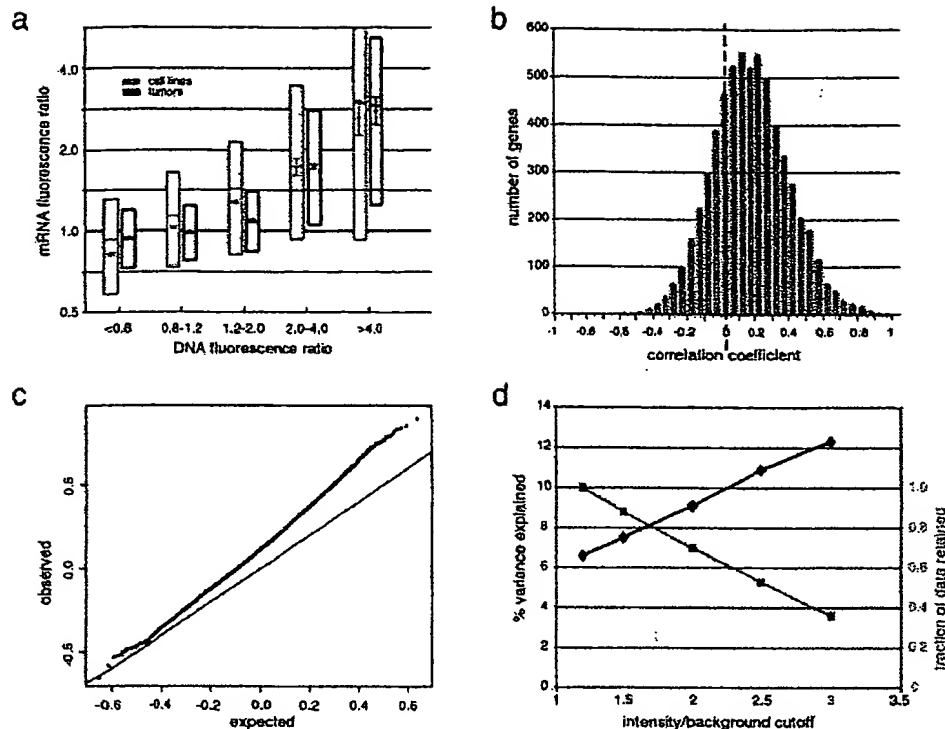


Fig. 3. Concordance between DNA copy number and gene expression across chromosome 17. DNA copy number alteration (Upper) and mRNA levels (Lower) are illustrated for breast cancer cell lines and tumors. Breast cancer cell lines and tumors are separately ordered by hierarchical clustering (Upper), and the identical sample order is maintained (Lower). The 354 genes present on the microarrays and mapping to chromosome 17, and for which both DNA copy number and mRNA levels were determined, are ordered by position along the chromosome; selected genes are indicated in color-coded text (see Fig. 2 legend). Fluorescence ratios (test/reference) are depicted by separate  $\log_2$  pseudocolor scales (indicated).

of DNA copy number and mRNA levels for genes on chromosome 17 (Fig. 3). The overall patterns of gene amplification and elevated gene expression are quite concordant; i.e., a significant fraction of highly amplified genes appear to be correspondingly highly expressed. The concordance between high-level amplification and increased gene expression is not restricted to chromosome 17. Genome-wide, of 117 high-level DNA amplifications (fluorescence ratios  $>4$ , and representing 91 different genes), 62% (representing 54 different genes; see Table 5, which is published as supporting information on the PNAS web site) are found associated with at least moderately elevated mRNA levels (mean-centered fluorescence ratios  $>2$ ), and 42% (representing 36 different genes) are found associated with comparably highly elevated mRNA levels (mean-centered fluorescence ratios  $>4$ ).

To determine the extent to which DNA deletion and lower-level amplification (in addition to high-level amplification) are also associated with corresponding alterations in mRNA levels, we performed three separate analyses on the complete data set (4 cell lines and 37 tumors, across 6,095 genes). First, we determined the average mRNA levels for each of five classes of genes, representing DNA deletion, no change, and low-, medium-, and high-level amplification (Fig. 4a). For both the

breast cancer cell lines and tumors, average mRNA levels tracked with DNA copy number across all five classes, in a statistically significant fashion ( $P$  values for pair-wise Student's  $t$  tests comparing adjacent classes: cell lines,  $4 \times 10^{-49}$ ,  $1 \times 10^{-49}$ ,  $5 \times 10^{-5}$ ,  $1 \times 10^{-2}$ ; tumors,  $1 \times 10^{-43}$ ,  $1 \times 10^{-214}$ ,  $5 \times 10^{-41}$ ,  $1 \times 10^{-4}$ ). A linear regression of the average  $\log(\text{DNA copy number})$ , for each class, against average  $\log(\text{mRNA level})$  demonstrated that on average, a 2-fold change in DNA copy number was accompanied by 1.4- and 1.5-fold changes in mRNA level for the breast cancer cell lines and tumors, respectively (Fig. 4a, regression line not shown). Second, we characterized the distribution of the 6,095 correlations between DNA copy number and mRNA level, each across the 37 tumor samples (Fig. 4b). The distribution of correlations forms a normal-shaped curve, but with the peak markedly shifted in the positive direction from zero. This shift is statistically significant, as evidenced in a plot of observed vs. expected correlations (Fig. 4c), and reflects a pervasive global influence of DNA copy number alterations on gene expression. Notably, the highest correlations between DNA copy number and mRNA level (the right tail of the distribution in Fig. 4b) comprise both amplified and deleted genes (data not shown). Third, we used a linear regression model to estimate the fraction of all variation measured in mRNA levels among the 37



**Fig. 4.** Genome-wide influence of DNA copy number alterations on mRNA levels. (a) For breast cancer cell lines (gray) and tumor samples (black), both mean-centered mRNA fluorescence ratio (log<sub>2</sub> scale) quartiles (box plots indicate 25th, 50th, and 75th percentile) and averages (diamonds; Y-value error bars indicate standard errors of the mean) are plotted for each of five classes of genes, representing DNA deletion (tumor/normal ratio < 0.8), no change (0.8–1.2), low- (1.2–2), medium- (2–4), and high-level (>4) amplification. *P* values for pair-wise Student's *t* tests, comparing averages between adjacent classes (moving left to right), are  $4 \times 10^{-49}$ ,  $1 \times 10^{-49}$ ,  $5 \times 10^{-5}$ ,  $1 \times 10^{-2}$  (cell lines), and  $1 \times 10^{-43}$ ,  $1 \times 10^{-214}$ ,  $5 \times 10^{-41}$ ,  $1 \times 10^{-4}$  (tumors). (b) Distribution of correlations between DNA copy number and mRNA levels, for 6,095 different human genes across 37 breast tumor samples. (c) Plot of observed versus expected correlation coefficients. The expected values were obtained by randomization of the sample labels in the DNA copy number data set. The line of unity is indicated. (d) Percent variance in gene expression (among tumors) directly explained by variation in gene copy number. Percent variance explained (black line) and fraction of data retained (gray line) are plotted for different fluorescence intensity/background (a rough surrogate for signal/noise) cutoff values. Fraction of data retained is relative to the 1.2 intensity/background cutoff. Details of the linear regression model used to estimate the fraction of variation in gene expression attributable to underlying DNA copy number alteration can be found in the supporting information (see *Estimating the Fraction of Variation in Gene Expression Attributable to Underlying DNA Copy Number Alteration*).

tumors that could be attributed to underlying variation in DNA copy number. From this analysis, we estimate that, overall, about 7% of all of the observed variation in mRNA levels can be explained directly by variation in copy number of the altered genes (Fig. 4d). We can reduce the effects of experimental measurement error on this estimate by using only that fraction of the data most reliably measured (fluorescence intensity/background >3); using that data, our estimate of the percent variation in mRNA levels directly attributed to variation in gene copy number increases to 12% (Fig. 4d). This still undoubtedly represents a significant underestimate, as the observed variation in global gene expression is affected not only by true variation in the expression programs of the tumor cells themselves, but also by the variable presence of non-tumor cell types within clinical samples.

## Discussion

This genome-wide, array CGH analysis of DNA copy number alteration in a series of human breast tumors demonstrates the usefulness of defining amplicon boundaries at high resolution (gene-by-gene), and quantitatively measuring amplicon shape, to assist in locating and identifying candidate oncogenes. By analyzing mRNA levels in parallel, we have also discovered that changes in DNA copy number have a large, pervasive, direct effect on global gene expression patterns in both breast cancer

cell lines and tumors. Although the DNA microarrays used in our analysis may display a bias toward characterized and/or highly expressed genes, because we are examining such a large fraction of the genome (approximately 20% of all human genes), and because, as detailed above, we are likely underestimating the contribution of DNA copy number changes to altered gene expression, we believe our findings are likely to be generalizable (but would nevertheless still be remarkable if only applicable to this set of ~6,100 genes).

In budding yeast, aneuploidy has been shown to result in chromosome-wide gene expression biases (13). Two recent studies have begun to examine the global relationship between DNA copy number and gene expression in cancer cells. In agreement with our findings, Phillips *et al.* (14) have shown that with the acquisition of tumorigenicity in an immortalized prostate epithelial cell line, new chromosomal gains and losses resulted in a statistically significant respective increase and decrease in the average expression level of involved genes. In contrast, Platzer *et al.* (15) recently reported that in metastatic colon tumors only ~4% of genes within amplified regions were found more highly (>2-fold) expressed, when compared with normal colonic epithelium. This report differs substantially from our finding that 62% of highly amplified genes in breast cancer exhibit at least 2-fold increased expression. These contrasting findings may reflect methodological differences between the

studies. For example, the study of Platzer *et al.* (15) may have systematically under-measured gene expression changes. In this regard it is remarkable that only 14 transcripts of many thousand residing within unamplified chromosomal regions were found to exhibit at least 4-fold altered expression in metastatic colon cancer. Additionally, their reliance on lower-resolution chromosomal CGH may have resulted in poorly delimiting the boundaries of high-complexity amplicons, effectively overcalling regions with amplification. Alternatively, the contrasting findings for amplified genes may represent real biological differences between breast and metastatic colon tumors; resolution of this issue will require further studies.

Our finding that widespread DNA copy number alteration has a large, pervasive and direct effect on global gene expression patterns in breast cancer has several important implications. First, this finding supports a high degree of copy number-dependent gene expression in tumors. Second, it suggests that most genes are not subject to specific autoregulation or dosage compensation. Third, this finding cautions that elevated expression of an amplified gene cannot alone be considered strong independent evidence of a candidate oncogene's role in tumorigenesis. In our study, fully 62% of highly amplified genes demonstrated moderately or highly elevated expression. This highlights the importance of high-resolution mapping of amplicon boundaries and shape [to identify the "driving" gene(s) within amplicons (16)], on a large number of samples, in addition to functional studies. Fourth, this finding suggests that analyzing

the genomic distribution of expressed genes, even within existing microarray gene expression data sets, may permit the inference of DNA copy number aberration, particularly aneuploidy (where gene expression can be averaged across large chromosomal regions; see Fig. 3 and supporting information). Fifth, this finding implies that a substantial portion of the phenotypic uniqueness (and by extension, the heterogeneity in clinical behavior) among patients' tumors may be traceable to underlying variation in DNA copy number. Sixth, this finding supports a possible role for widespread DNA copy number alteration in tumorigenesis (17, 18), beyond the amplification of specific oncogenes and deletion of specific tumor suppressor genes. Widespread DNA copy number alteration, and the concomitant widespread imbalance in gene expression, might disrupt critical stoichiometric relationships in cell metabolism and physiology (e.g., proteasome, mitotic spindle), possibly promoting further chromosomal instability and directly contributing to tumor development or progression. Finally, our findings suggest the possibility of cancer therapies that exploit specific or global imbalances in gene expression in cancer.

We thank the many members of the P.O.B. and D.B. labs for helpful discussions. J.R.P. was a Howard Hughes Medical Institute Physician Postdoctoral Fellow during a portion of this work. P.O.B. is a Howard Hughes Medical Institute Associate Investigator. This work was supported by grants from the National Institutes of Health, the Howard Hughes Medical Institute, the Norwegian Cancer Society, and the Norwegian Research Council.

- Kallioniemi, A., Kallioniemi, O. P., Sudar, D., Rutovitz, D., Gray, J. W., Waldman, F. & Pinkel, D. (1992) *Science* 258, 818–821.
- Kallioniemi, A., Kallioniemi, O. P., Piper, J., Tanner, M., Stokke, T., Chen, L., Smith, H. S., Pinkel, D., Gray, J. W. & Waldman, F. M. (1994) *Proc. Natl. Acad. Sci. USA* 91, 2156–2160.
- Tirkkonen, M., Tanner, M., Karhu, R., Kallioniemi, A., Isola, J. & Kallioniemi, O. P. (1998) *Genes Chromosomes Cancer* 21, 177–184.
- Forozan, F., Mahlamaki, E. H., Monni, O., Chen, Y., Veldman, R., Jiang, Y., Gooden, G. C., Ethier, S. P., Kallioniemi, A. & Kallioniemi, O. P. (2000) *Cancer Res.* 60, 4519–4525.
- Solinas-Toldo, S., Lampel, S., Stigenbauer, S., Nickolenko, J., Benner, A., Dohner, H., Cremer, T. & Lichter, P. (1997) *Genes Chromosomes Cancer* 20, 399–407.
- Pinkel, D., Seagraves, R., Sudar, D., Clark, S., Poole, I., Kowbel, D., Collins, C., Kuo, W. L., Chen, C., Zhai, Y., *et al.* (1998) *Nat. Genet.* 20, 207–211.
- Pollack, J. R., Perou, C. M., Alizadeh, A. A., Eisen, M. B., Pergamenschikov, A., Williams, C. F., Jeffrey, S. S., Botstein, D. & Brown, P. O. (1999) *Nat. Genet.* 23, 41–46.
- Perou, C. M., Sorlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, L. A., *et al.* (2000) *Nature (London)* 406, 747–752.
- Sorlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., *et al.* (2001) *Proc. Natl. Acad. Sci. USA* 98, 10869–10874.
- Schuler, G. D. (1997) *J. Mol. Med.* 75, 694–698.
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* (2001) *Nature (London)* 409, 860–921.
- Fejzo, M. S., Godfrey, T., Chen, C., Waldman, F. & Gray, J. W. (1998) *Genes Chromosomes Cancer* 22, 105–113.
- Hughes, T. R., Roberts, C. J., Dai, H., Jones, A. R., Meyer, M. R., Slade, D., Burchard, J., Dow, S., Ward, T. R., Kidd, M. J., *et al.* (2000) *Nat. Genet.* 25, 333–337.
- Phillips, J. L., Hayward, S. W., Wang, Y., Vasselli, J., Pavlovich, C., Padilla-Nash, H., Pezullo, J. R., Ghadimi, B. M., Grossfeld, G. D., Rivera, A., *et al.* (2001) *Cancer Res.* 61, 8143–8149.
- Platzer, P., Upender, M. B., Wilson, K., Willis, J., Lutterbaugh, J., Nosrati, A., Willson, J. K., Mack, D., Ried, T. & Markowitz, S. (2002) *Cancer Res.* 62, 1134–1138.
- Albertson, D. G., Ylstra, B., Seagraves, R., Collins, C., Dairkee, S. H., Kowbel, D., Kuo, W. L., Gray, J. W. & Pinkel, D. (2000) *Nat. Genet.* 25, 144–146.
- Li, R., Yerganian, G., Duesberg, P., Kraemer, A., Willer, A., Rausch, C. & Hehlmann, R. (1997) *Proc. Natl. Acad. Sci. USA* 94, 14506–14511.
- Rasnick, D. & Duesberg, P. H. (1999) *Biochem. J.* 340, 621–630.

## Variable expression of the translocated *c-abl* oncogene in Philadelphia-chromosome-positive B-lymphoid cell lines from chronic myelogenous leukemia patients

JAMES B. KONOPKA<sup>\*,‡</sup>, STEVEN CLARK<sup>\*</sup>, JAMI McLAUGHLIN<sup>\*</sup>, MASAKUZU NITTA<sup>†</sup>, YOSHIRO KATO<sup>†</sup>, ANNABEL STRIFE<sup>†</sup>, BAYARD CLARKSON<sup>†</sup>, AND OWEN N. WITTE<sup>\*,§</sup>

<sup>\*</sup>Department of Microbiology and Molecular Biology Institute, University of California, Los Angeles, 405 Hilgard Avenue, Los Angeles, CA 90024; and <sup>†</sup>The Laboratory of Hematopoietic Cell Kinetics and The Laboratory of Cancer Genetics and Cytogenetics, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, NY 10021

Communicated by Michael Potter, February 10, 1986

**ABSTRACT** The consistent cytogenetic translocation of chronic myelogenous leukemia (the Philadelphia chromosome, Ph<sup>1</sup>) has been observed in cells of multiple hematopoietic lineages. This translocation creates a chimeric gene composed of breakpoint-cluster-region (*bcr*) sequences from chromosome 22 fused to a portion of the *abl* oncogene on chromosome 9. The resulting gene product (P210<sup>c-abl</sup>) resembles the transforming protein of the Abelson murine leukemia virus in its structure and tyrosine kinase activity. P210<sup>c-abl</sup> is expressed in Ph<sup>1</sup>-positive cell lines of myeloid lineage and in clinical specimens with myeloid predominance. We show here that Epstein-Barr virus-transformed B-lymphocyte lines that retain Ph<sup>1</sup> can express P210<sup>c-abl</sup>. The level of expression in these B-cell lines is generally lower and more variable than that observed for myeloid lines. Protein expression is not related to amplification of the *abl* gene but to variation in the level of *bcr-abl* mRNA produced from a single Ph<sup>1</sup> template.

Chronic myelogenous leukemia (CML) is a disease of the pluripotent stem cell (1). In greater than 95% of patients, the leukemic cells contain the cytogenetic marker known as the Philadelphia chromosome, or Ph<sup>1</sup> (2). This reciprocal translocation event between the long arms of chromosomes 9 and 22 has been used as a disease-specific marker for diagnosis and evaluation of therapy. Multiple hematopoietic lineages, including myeloid and B-lymphoid, contain Ph<sup>1</sup> in early or chronic phase, as well as in the more acute accelerated and blast crisis phases of the disease.

One molecular consequence of Ph<sup>1</sup> is the translocation of the chromosomal arm containing the *c-abl* gene on chromosome 9 into the middle of the breakpoint-cluster region (*bcr*) gene on chromosome 22 (3-6). Although the precise translocation breakpoints are variable, an RNA-splicing mechanism generates a very similar 8-kilobase (kb) mRNA in each case (5-9). The hybrid *bcr-abl* message encodes a structurally altered form of the *abl* oncogene product, called P210<sup>c-abl</sup> (10-13), with an amino-terminal segment derived from a portion of the exons of *bcr* on chromosome 22 and a carboxyl-terminal segment derived from a major portion of the exons of the *c-abl* gene on chromosome 9. The chimeric structure of *bcr-abl* and the resulting P210<sup>c-abl</sup> is similar to the structure of the Abelson murine leukemia virus *gag-abl* genome and resulting P160<sup>v-abl</sup> transforming gene product. Both proteins have very similar tyrosine kinase activities (10, 11, 14) which can be distinguished by their relative stability to denaturing detergents and by their ATP requirements from the recently described tyrosine kinase activity of the *c-abl* gene product (15).

In concert with structural modification of the amino-terminal portion of the *abl* gene, increased level of expression has been implicated in activation of *c-abl* oncogenic potential. Myeloid and erythroid cell lines and clinical samples derived from acute-phase CML patients contain about 10-fold higher levels of the 8-kb *bcr-abl* mRNA and P210<sup>c-abl</sup> than the *c-abl* mRNA forms (6 and 7 kb) and P145<sup>c-abl</sup> gene product (5, 8, 9, 11). The higher level of expression of the chimeric *bcr-abl* message in acute-phase cells is not likely to be solely due to the presence of the *bcr* promoter sequences at the 5' end of the gene, since the normal 4.5-kb and 6.7-kb *bcr*-encoded mRNA species are expressed at an even lower level than the normal *c-abl* messages (5, 6).

We have analyzed a series of Epstein-Barr virus-immortalized B-lymphoid cell lines derived from CML patients (16). With such *in vitro* clonal cell lines, we can evaluate whether the presence of Ph<sup>1</sup> always results in synthesis of the chimeric *bcr-abl* message and protein, and whether the quantitative expression varies for cells of B-lymphoid lineage as compared to previously examined myeloid cell lines. Our results show that cell lines that retain Ph<sup>1</sup> do express *bcr-abl* message and protein, but that the level is generally lower and more variable than previously seen for myeloid cell lines. The demonstration that the Ph<sup>1</sup> chromosomal template can vary in its level of expression of P210<sup>c-abl</sup> suggests that secondary mechanisms, beyond the translocation itself, contribute to the regulation of the *bcr-abl* gene in different cell types or subclones that derive from the affected stem cell.

### MATERIALS AND METHODS

**Cells and Cell Labelings.** Epstein-Barr virus-transformed B-lymphoid cell lines were established from peripheral blood samples of chronic- and acute-phase CML patients as reported (16). The cell lines are designated according to patient number, karyotype, and lineage. For example, SK-CML7Bt(9,22)-33 refers to CML patient 7, B-lymphoid cell line, 9;22 translocation (Ph<sup>1</sup>), cell line 33; and SK-CML7BN-2 refers to B-cell line 2 with a normal karyotype derived from the same patient. Repeat karyotype analysis was performed to verify the retention of Ph<sup>1</sup> just prior to analysis for *abl* protein and RNA. Cells were maintained in RPMI 1640 medium with 20% fetal bovine serum. We have not observed any consistent pattern of *in vitro* growth rate that correlates to the stage of disease at the time of transformation with Epstein-Barr virus. Cells ( $1.5 \times 10^7$ ) were washed twice with Dulbecco's modified Eagle's medium lacking phosphate and

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: *bcr*, breakpoint-cluster region; CML, chronic myelogenous leukemia; kb, kilobase(s).

<sup>‡</sup>Present address: Department of Genetics, University of Washington, Seattle, WA 98195.

<sup>§</sup>To whom correspondence should be addressed.

supplemented with 5% dialyzed fetal bovine serum. Cells were then resuspended in 2 ml of the minimal medium. Labeling was started with the addition of [ $^{32}$ P]orthophosphate (1 mCi/ml; ICN; 1 Ci = 37 GBq) and continued at 37°C for 3–4 hr.

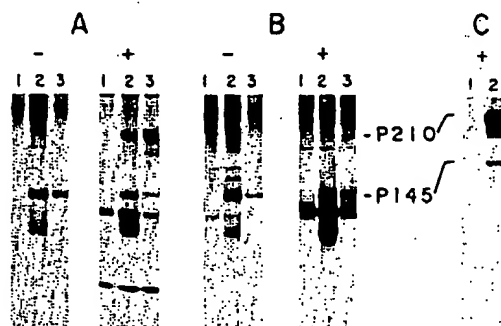
**Immunoprecipitation and Immunoblotting.** Immunoprecipitations were carried out as described (10). Cells ( $1.5 \times 10^7$ ) were washed with phosphate-buffered saline and extracted with 3–5 ml of phosphate lysis buffer (1% Triton X-100/0.1 NaDodSO<sub>4</sub>/0.5% deoxycholate/10 mM Na<sub>2</sub>HPO<sub>4</sub>, pH 7.5/100 mM NaCl) with 5 mM EDTA and 5 mM phenylmethylsulfonyl fluoride. Extracts were clarified by centrifugation and precipitated with normal or rabbit anti-abl sera (anti-pEX-2 or anti-pEX-5) (17). The precipitated proteins were electrophoresed in a NaDodSO<sub>4</sub>/8% polyacrylamide gel.  $^{32}$ P-labeled proteins were detected by autoradiography. Alternatively, *abl* proteins were detected by immunoblotting. Extracts from unlabeled cells were clarified, and proteins were concentrated by immunoprecipitation with rabbit antisera against *abl*-encoded proteins [anti-pEX-2 and anti-pEX-5 combined (17)] and then fractionated in 8% acrylamide gels. The proteins were transferred from the gel to nitrocellulose filters, using protease-facilitated transfer (18). The *abl*-encoded proteins were detected using murine monoclonal antibodies as a probe and peroxidase-conjugated goat anti-mouse second stage antibody (Bio-Rad) for development. Rabbit antisera and mouse monoclonal antibodies to *abl* proteins were prepared using bacterially expressed regions of the *v-abl* protein as immunogens (17, 19). Anti-pEX-2 antibodies react with the internal tyrosine kinase domain and anti-pEX-5 antibodies react with the carboxyl-terminal segment of the *abl* proteins.

**RNA Analysis.** RNA was extracted from  $10^8$  cells by the NaDodSO<sub>4</sub>/urea/phenol method (20). Polyadenylated RNA was purified by oligo(dT) affinity chromatography. Samples were electrophoresed in a 1% agarose/formaldehyde gel and transferred to nitrocellulose. *abl* RNA species were detected by hybridization with a nick-translated *v-abl* fragment probe (21).

**DNA Analysis.** DNA was prepared from  $5 \times 10^7$  cells of each cell line and processed for Southern blots with a *v-abl* probe as described (21).

## RESULTS

**Variable Levels of P210<sup>c-abl</sup> Are Detected in Ph<sup>1</sup>-Positive Cell Lines.** Ph<sup>1</sup>-positive and Ph<sup>1</sup>-negative, Epstein-Barr virus-transformed B-lymphocyte cell lines derived from the same patient were examined for P210<sup>c-abl</sup> synthesis by immunoprecipitation of [ $^{32}$ P]orthophosphate-labeled cell extracts with anti-abl sera (Fig. 1). The normal *c-abl* protein P145<sup>c-abl</sup> was detected at a similar level in multiple Ph<sup>1</sup>-positive and Ph<sup>1</sup>-negative cell lines. P210<sup>c-abl</sup> was only detected in the Ph<sup>1</sup>-positive cell lines because the *bcr-abl* chimeric gene which encodes P210<sup>c-abl</sup> resides on the Ph<sup>1</sup> (4, 5, 11, 13). The level of P210<sup>c-abl</sup> was about 4- to 5-fold higher than the level of P145<sup>c-abl</sup> in the SK-CML7Bt-33 cell line (Fig. 1A, +). The Ph<sup>1</sup>-positive erythroid-progenitor cell line K562 (C) showed a level of P210<sup>c-abl</sup> about 10-fold higher than P145<sup>c-abl</sup>. However, the level of P210<sup>c-abl</sup> was about one-fifth that of P145<sup>c-abl</sup> in the Ph<sup>1</sup>-positive SK-CML16Bt-1 cell line (Fig. 1B, +). Comparison of different autoradiographic exposures roughly indicated that the level of P210<sup>c-abl</sup> varies over a 20-fold range between these Ph<sup>1</sup>-positive B-cell lines. Analysis of four additional Ph<sup>1</sup>-positive B-cell lines demonstrated that the level of P210<sup>c-abl</sup> fell into two general classes; some cell lines had a level of P210<sup>c-abl</sup> similar to SK-CML7Bt-33 and others had the low level similar to SK-CML16Bt-1 (Table 1). This differs from previous studies with Ph<sup>1</sup>-positive myeloid cell lines and patient samples derived from acute-



**FIG. 1.** Detection of variable levels of P210<sup>c-abl</sup> in Ph<sup>1</sup>-positive B-cell lines. Production of P145<sup>c-abl</sup> and P210<sup>c-abl</sup> in Epstein-Barr virus-transformed B-cell lines derived from a blast-crisis (A) and a chronic-phase (B) CML patient was examined by metabolic labeling with [ $^{32}$ P]orthophosphate and immunoprecipitation. Ph<sup>1</sup>-negative (–) and Ph<sup>1</sup>-positive (+) cell lines derived from each patient were analyzed. The Ph<sup>1</sup>-negative cell line in A, – is SK-CML7BN-2 and in B, – is SK-CML16BN-1. The Ph<sup>1</sup>-positive cell line in A, + is SK-CML7Bt-33 and in B, + is SK-CML16Bt-1. The K562 cell line, a Ph<sup>1</sup>-positive erythroid progenitor cell line spontaneously derived from a blast-crisis patient (33), is represented in C. Cells ( $1.5 \times 10^7$ ) were metabolically labeled with 2 mCi of [ $^{32}$ P]orthophosphate for 3–4 hr and then were extracted and clarified by centrifugation. Samples were immunoprecipitated with control normal serum (lanes 1), anti-pEX-2 (lanes 2), or anti-pEX-5 (lanes 3) and analyzed by NaDodSO<sub>4</sub>/8% PAGE followed by autoradiography with an intensifying screen (3 days for A and C, 10 days for B).

phase CML patients, in which P210<sup>c-abl</sup> was detected at a 10-fold higher level than P145<sup>c-abl</sup> (refs. 10 and 11; Table 1). There was no large difference in level of chimeric mRNA and P210<sup>c-abl</sup> expressed in four myeloid/erythroid-lineage Ph<sup>1</sup>-positive cell lines (K562, EM2, EM3, CML22, and BV173; refs. 9 and 11), despite a 4- to 5-fold amplification of *abl*-related sequences in the K562 cell line.

Detection of different levels of P210<sup>c-abl</sup> in Fig. 1 could be due to decreased phosphorylation of P210<sup>c-abl</sup>, a lower level of P210<sup>c-abl</sup> synthesis, or altered stability of the protein. To help distinguish among these possibilities, the steady-state level of P210<sup>c-abl</sup> in the cell lines was assayed by immunoblotting. The results show that SK-CML7Bt-33 (Fig. 2A, +) had a higher level of P210<sup>c-abl</sup> than P145, similar to the results with metabolic labeling (Fig. 1). We did not detect P210<sup>c-abl</sup> by immunoblotting with  $2 \times 10^7$  cells of line SK-CML8Bt-3 (Fig. 2B, +). Reconstruction experiments using dilutions of cell extracts showed that we could detect about 5–10% the level of P210<sup>c-abl</sup> expressed in the K562 cell line (data not shown). We infer that the steady-state level of P210<sup>c-abl</sup> in SK-CML8Bt-3 is lower than the level in SK-CML7Bt-33 by a factor of at least 10. The level of P210<sup>c-abl</sup> detected in these assays correlated with the amount of P210<sup>c-abl</sup> tyrosine kinase activity that could be detected *in vitro* (data not shown).

**Different Levels of P210<sup>c-abl</sup> Are Reflected in the Amount of Stable *bcr-abl* mRNA.** To identify the basis for detection of variable levels of P210<sup>c-abl</sup>, we examined the production of the *abl* RNA. RNA blot hybridization analysis using a *v-abl* probe (Fig. 3) showed that the normal 6- and 7-kb *c-abl* mRNAs were present at a similar level in Ph<sup>1</sup>-positive and -negative cell lines derived from different patients. However, the 8-kb mRNA that encodes P210<sup>c-abl</sup> was detected at a 10-fold higher level in SK-CML7Bt-33 (Fig. 3A, +) than in SK-CML16Bt-1 (B, +), which correlated with the relative level of P210<sup>c-abl</sup> detected in each cell line. Analysis of additional cell lines demonstrated that the level of 8-kb RNA directly correlated with the level of P210<sup>c-abl</sup> (Table 1). The variation in level of 8-kb RNA detected in these cell lines was not due to loss or gain of Ph<sup>1</sup>, because cytogenetic analysis confirmed the presence of Ph<sup>1</sup> in these cell lines (ref. 16 and

Table 1. Relative levels of *bcr-abl* expression in Epstein-Barr virus-immortalized B-cell lines and myeloid CML lines

Cell line*	CML phase†	Ph <sup>1</sup> ‡	P210§	8-kb mRNA¶
SK-CML7BN-2	BC	-	-	-
SK-CML8BN-10	Chronic	-	-	-
SK-CML8BN-12	Chronic	-	-	-
SK-CML16BN-1	Chronic	-	-	-
SK-CML35BN-1	Chronic	-	-	-
SK-CML7B5-33	BC	+	+++	+++
SK-CML21Bt-1	Acc	+	+++	+++
SK-CML21Bt-6	Acc	+	+++	+++
SK-CML8Bt-3	Chronic	+	+	±
SK-CML16Bt-1	Chronic	+	+	+
SK-CML35Bt-2	Chronic	+	+	+
K562	BC	+	+++++	+++++
BV173	BC	+	+++++	+++++
EM2	BC	+	+++++	+++++

\*Cell lines derived from CML patients by transformation with Epstein-Barr virus as described (16). Names of cell lines indicate patient number and Ph<sup>1</sup> status: SK-CML7Bt indicates a cell line derived from patient 7 that carries the 9;22 Ph<sup>1</sup> translocation; N indicates a normal karyotype. Myeloid-erythroid cell lines (K562, EM2, and BV173) are described in previous publications (9, 11, 22, 33).

†Status of patient at the time cell line was derived. BC, blast crisis; Acc, accelerated phase.

‡Presence (+) or absence (-) of Ph<sup>1</sup> as demonstrated by karyotypic or Southern blot analysis.

§P210<sup>c-abl</sup> detected as described in legend to Fig. 1. B-cell lines derived from blast-crisis and accelerated-phase patients had levels of P210 3- to 5-fold higher (++++) than levels of P145. Chronic-phase-derived cell lines had P210 levels lower than or just equivalent (+) to the level of P145. Myeloid and erythroid lines had levels of P210 5- to 10-fold higher than P145 (+++++).

¶Eight-kilobase *bcr-abl* mRNA detected as described in legend to Fig. 2. Symbols: ±, borderline detectable; +++++, level of 8-kb mRNA 5- to 10-fold higher than that of the 6- and 7-kb *c-abl* mRNA species; +++, level of 8-kb mRNA 3- to 5-fold higher than that of the 6- and 7-kb messages; +, a level approximately equivalent to that of the 6- and 7-kb messages.

data not shown). There was no difference in the copy number of *abl*-related sequences as judged by Southern blot analysis (Fig. 4). Only the K562 cell line control showed an amplification of *abl* sequences, as previously reported (22, 23). These combined data suggest that differential *bcr-abl* mRNA expression from a single gene template is responsible for the variable levels of P210<sup>c-abl</sup> detected. This could be mediated

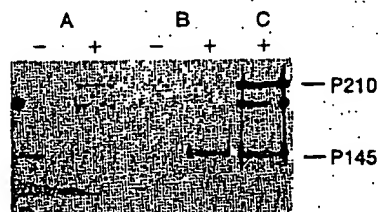


FIG. 2. Analysis of steady-state *abl* protein levels by immunoblotting. Cell extracts prepared from  $2 \times 10^7$  cells of lines SK-CML7BN-2 (A, -), SK-CML7Bt-33 (A, +), SK-CML8BN-10 (B, -), and SK-CML8Bt-3 (B, +) were concentrated by immunoprecipitation with anti-pEX-2 plus anti-pEX-5. Samples were then electrophoresed in a NaDodSO<sub>4</sub>/8% polyacrylamide gel and transferred to nitrocellulose, using protease-facilitated transfer (18). *abl* proteins were detected using a mixture of two monoclonal antibodies directed against the pEX-2 and pEX-5 *abl*-protein fragments produced in bacteria (19) as a probe and a peroxidase-conjugated goat anti-mouse second-stage antibody (Bio-Rad) for development.

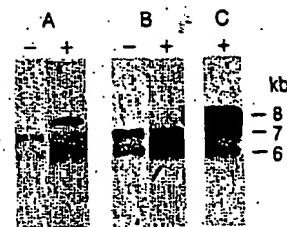


FIG. 3. Comparison of *abl* RNA levels in Ph<sup>1</sup>-positive and -negative B-cell lines. The levels of the normal 6- and 7-kb *c-abl* RNAs and the 8-kb *bcr-abl* RNA were analyzed by blot hybridization using a v-*abl* probe. RNA was extracted from Ph<sup>1</sup>-negative lines SK-CML7BN-2 (A, -) and SK-CML16BN-1 (B, -), from Ph<sup>1</sup>-positive lines SK-CML6Bt-33 (A, +) and SK-CML16Bt-3 (B, +), and from line K562 (C, +) by the NaDodSO<sub>4</sub>/urea/phenol method (20). Polyadenylated RNA was purified by oligo(dT) affinity chromatography, and 15 µg of each sample was electrophoresed in a 1% agarose/formaldehyde gel and then transferred to nitrocellulose. The blotted RNAs were hybridized with a nick-translated v-*abl* fragment probe (21) and then autoradiographed for 4 days.

by factors influencing the transcription rate of the *bcr-abl* gene or the stability of the mRNA.

## DISCUSSION

Several lines of evidence suggest that formation of Ph<sup>1</sup> is not the primary event that affects the stem cell in CML. Patients have been identified that present with the clinical picture of CML but only later develop Ph<sup>1</sup> (1). This observation, coupled with studies of *G6PD* (glucose-6-phosphate dehydrogenase)-heterozygous females with CML that demonstrate stem-cell clonality by isozyme analysis among cell

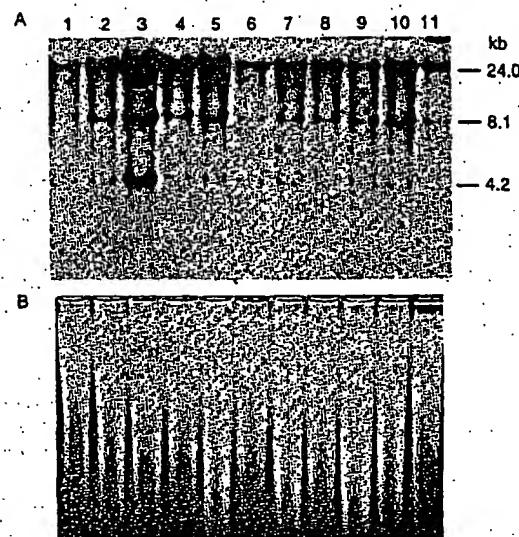


FIG. 4. Southern blot analysis of *abl* sequences in Ph<sup>1</sup>-positive and -negative B-cell lines. High molecular weight DNA (15 µg) was digested with restriction endonuclease *Bam*HI, separated in a 0.8% agarose gel, and then transferred to nitrocellulose. The blotted DNA fragments were hybridized with a nick-translated, 2.4-kb *Bgl* II v-*abl* fragment ( $1.5 \times 10^8$  cpm/µg; ref. 21) and exposed for 4 days. (A) Autoradiogram of *abl*-specific fragments in cell lines HL-60 (lane 1), EM2 (lane 2), K562 (lane 3), SK-CML7Bt-33 (lane 4), SK-CML8Bt-3 (lane 5), SK-CML16Bt-1 (lane 6), SK-CML21Bt-6 (lane 7), SK-CML35Bt-2 (lane 8), SK-CML7BN-2 (lane 9), SK-CML8BN-2 (lane 10), and SK-CML35BN-1 (lane 11). (B) Ethidium bromide staining of agarose gel prior to transfer to nitrocellulose, showing the level of variation in amount of DNA loaded per lane.

populations that lack the Ph<sup>1</sup> marker, supports a secondary or complementary role for Ph<sup>1</sup> in the progression of the disease (24, 25). This chromosome marker is found in chronic, accelerated, and blast-crisis phases of the disease. It is likely that Ph<sup>1</sup> confers some growth advantage, since cells with the marker chromosome eventually predominate the marrow and peripheral blood even in chronic phase. During the phase of blast crisis, many patients develop additional chromosome abnormalities, including duplication of Ph<sup>1</sup>, a variety of trisomies, and complex translocations (26). This is suggestive evidence for Ph<sup>1</sup> being a necessary but not sufficient genetic change for the full evolution of the disease.

The realization that one molecular result of Ph<sup>1</sup> is the generation of a chimeric *bcr-abl* protein with functional characteristics and structure analogous to the *gag-abl* transforming protein of the Abelson murine leukemia virus strengthens the argument for an important role of Ph<sup>1</sup> in the pathogenesis of CML. Although the Abelson virus is generally considered a rapidly transforming retrovirus, its effects can range from overcoming growth factor requirements, to cellular lethality, to induction of highly oncogenic tumors in a number of hematopoietic cell lineages (27, 28). Even in the transformation of murine cell targets, there are several lines of evidence that suggest that the growth-promoting activity of the *v-abl* gene product is complemented by further cellular changes in the production of the malignant-cell phenotype (29–31).

The regulation of *bcr-abl* gene expression is complex because the 5' end of the gene is derived from the non-*abl* sequences, *bcr*, normally found on chromosome 22 (6). The level of stable message for the normal *bcr* gene and the normal *abl* gene are both much lower than the level of the *bcr-abl* message and protein from cell lines and clinical specimens derived from myeloid blast-crisis patients (5, 6, 11). Therefore, the high level of *bcr-abl* expression cannot simply be attributed to the regulatory sequences associated with *bcr*. Possibly, creation of the chimeric gene disrupts the normal regulatory sequences and results in a higher level of expression. Variation in *bcr-abl* expression may result from secondary changes in the structure of the chimeric gene or function of *trans*-acting factors that occur during evolution of the disease. Our analysis of P210<sup>c-abl</sup> and the 8-kb mRNA in Epstein-Barr virus-transformed Ph<sup>1</sup>-positive B-cell lines demonstrates that stable message and protein levels from the *bcr-abl* gene can vary over a wide range. This variation does not result from a change in the number of *bcr-abl* templates secondary to gene amplification but more likely from changes in either transcription rate or mRNA stability. We suspect this range of *bcr-abl* expression is not limited to lymphoid cells. Analysis of peripheral blood leukocytes derived from an unusual CML patient who has been in chronic phase with myeloid predominance for 16 years showed a level of P210<sup>c-abl</sup> one-fifth that of P145<sup>c-abl</sup>, as detected by metabolic labeling with [<sup>32</sup>P]orthophosphate and immunoprecipitation (S.C., O.N.W., and P. Greenberg, unpublished observations). Lower levels of expression of the chimeric mRNA have been demonstrated in clinical samples from chronic-phase CML patients compared to acute-phase CML patients (9). Others have reported chronic-phase patients with variable but, in some cases, relatively high levels of the *bcr-abl* mRNA (32). The sampling variation and the heterogeneous mixture of cell types in clinical samples complicate such analyses. Further work is needed to evaluate whether there is a defined change in P210<sup>c-abl</sup> expression during the progression of CML. It is interesting to note that among the limited sample of Ph<sup>1</sup>-positive B-cell lines we have examined (Table 1), we have seen higher levels of P210<sup>c-abl</sup> in those derived from patients at more advanced stages of the disease.

It will be important to search for cell-type-specific mechanisms that might regulate expression of *bcr-abl* from Ph<sup>1</sup>.

We thank Bonnie Hechinger and Carol Crookshank for excellent secretarial assistance and Margaret Newman for excellent technical assistance. This work was supported by grants from the National Institutes of Health (to O.N.W. and B.C.). J.B.K. was supported as a predoctoral fellow on the Public Health Service Cellular and Molecular Biology Training Grant GM07185. S.C. is a postdoctoral fellow of the Leukemia Society of America.

- Champlin, R. E. & Golde, D. W. (1985) *Blood* 65, 1039–1047.
- Rowley, J. D. (1973) *Nature (London)* 243, 290–291.
- Heisterkamp, N., Stephenson, J. R., Groffen, J., Hansen, P. F., de Klein, A., Bartram, C. R. & Grosfeld, G. (1983) *Nature (London)* 306, 239–242.
- Bartram, C. R., de Klein, A., Hagemeijer, A., van Agthoven, T., van Kessel, A. G., Bootsma, D., Grosfeld, G., Ferguson-Smith, M. A., Davies, T., Stone, M., Heisterkamp, N., Stephenson, J. R. & Groffen, J. (1983) *Nature (London)* 306, 277–280.
- Shtivelman, E., Lifshitz, B., Gale, R. P. & Canaani, D. (1985) *Nature (London)* 315, 550–554.
- Heisterkamp, N., Stam, K. & Groffen, J. (1985) *Nature (London)* 315, 758–761.
- Groffen, J., Stephenson, J. R., Heisterkamp, N., de Klein, A., Bartram, C. R. & Grosfeld, G. (1984) *Cell* 36, 93–99.
- Gale, R. P. & Canaani, E. (1984) *Proc. Natl. Acad. Sci. USA* 81, 5648–5652.
- Collins, S., Kubonishi, L., Miyoshi, I. & Groudine, M. T. (1984) *Science* 225, 72–74.
- Konopka, J. B., Watanabe, S. M. & Witte, O. N. (1984) *Cell* 7, 1035–1042.
- Konopka, J. B., Watanabe, S. M., Singer, J., Collins, S. & Witte, O. N. (1985) *Proc. Natl. Acad. Sci. USA* 82, 1810–1814.
- Kloetzer, W., Kurzrock, R., Smith, L., Talpaz, M., Spiller, M., Gutterman, J. & Arlinghaus, R. (1985) *Virology* 140, 230–238.
- Kozbor, D., Giallongo, A., Sierzega, M. E., Konopka, J. B., Witte, O. N., Showe, L. C. & Croce, C. M. (1985) *Nature (London)*, in press.
- Davis, R. L., Konopka, J. B. & Witte, O. N. (1985) *Mol. Cell Biol.* 5, 204–213.
- Konopka, J. B. & Witte, O. N. (1985) *Mol. Cell Biol.* 5, 3116–3123.
- Nitta, M., Kato, Y., Strife, A., Wachter, M., Fried, J., Perez, A., Jhanwar, S., Duigou, R., Chaganti, R. S. K. & Clarkson, B. (1985) *Blood* 66, 1053–1061.
- Konopka, J. B., Davis, J. L., Watanabe, S. M., Ponticelli, A. S., Schiff-Maker, L., Rosenberg, N. & Witte, O. N. (1984) *Virology* 51, 223–232.
- Gibson, W. (1981) *Anal. Biochem.* 118, 1–3.
- Schiff-Maker, L., Konopka, J. B., Clark, S., Witte, O. N. & Rosenberg, N. (1986) *J. Virol.* 57, 1182–1186.
- Schwartz, R. C., Sonenshein, G. E., Bothwell, A. & Gelfand, M. L. (1981) *J. Immunol.* 126, 2104–2108.
- Goff, S. P., Gilboa, E., Witte, O. N. & Baltimore, D. (1980) *Cell* 22, 777–785.
- Collins, S. J. & Groudine, M. T. (1983) *Proc. Natl. Acad. Sci. USA* 80, 4813–4817.
- Selden, J. R., Emanuel, B. S., Wang, E., Cannizzaro, L., Palumbo, A., Erikson, J., Nowell, P. C., Rovera, G. & Croce, C. M. (1983) *Proc. Natl. Acad. Sci. USA* 80, 7289–7292.
- Fialkow, P. J., Martin, P. J., Najfeld, V., Penfold, G. K., Jacobson, R. J. & Hansen, J. A. (1981) *Blood* 58, 158–163.
- Martin, P. J., Najfeld, V. & Fialkow, P. J. (1982) *Can. Gen. Cytogenet.* 6, 359–368.
- Rowley, J. D. (1980) *Annu. Rev. Genet.* 14, 17–40.
- Whitlock, C. A. & Witte, O. N. (1984) *Adv. Immunol.* 37, 74–98.
- Pierce, J. H., Di Fiore, P. P., Aaronson, S. A., Potter, M., Pumphrey, J., Scott, A. & Ihle, J. N. (1985) *Cell* 41, 685–693.
- Whitlock, C. A., Ziegler, S. & Witte, O. N. (1983) *Mol. Cell Biol.* 3, 596–604.
- Wolf, D., Harris, N. & Rotter, V. (1984) *Cell* 38, 119–126.
- Klein, G. & Klein, E. (1985) *Nature (London)* 315, 190–195.
- Stam, K., Jr., Heisterkamp, N., Grosfeld, G., de Klein, A., Verma, R., Coleman, M., Dosik, H. & Groffen, J. (1985) *N. Engl. J. Med.* 313, 1429–1433.
- Lozzio, C. B. & Lozzio, B. B. (1975) *Blood* 45, 321–334.

## WISP genes are members of the connective tissue growth factor family that are up-regulated in Wnt-1-transformed cells and aberrantly expressed in human colon tumors

DIANE PENNICA\*†, TODD A. SWANSON\*, JAMES W. WELSH\*, MARGARET A. ROY‡, DAVID A. LAWRENCE\*, JAMES LEE‡, JENNIFER BRUSH‡, LISA A. TANEYHILL§, BETHANNE DEUEL‡, MICHAEL LEW¶, COLIN WATANABE||, ROBERT L. COHEN\*, MONA F. MELHEM\*\*, GENE G. FINLEY\*\*, PHIL QUIRKE††, AUDREY D. GODDARD‡, KENNETH J. HILLAN¶, AUSTIN L. GURNEY‡, DAVID BOTSTEIN‡,†‡, AND ARNOLD J. LEVINE§

Departments of \*Molecular Oncology, ‡Molecular Biology, §Scientific Computing, and ¶Pathology, Genentech Inc., 1 DNA Way, South San Francisco, CA 94080; \*\*University of Pittsburgh School of Medicine, Veterans Administration Medical Center, Pittsburgh, PA 15240; ††University of Leeds, Leeds, LS29JT United Kingdom; ‡‡Department of Genetics, Stanford University, Palo Alto, CA 94305; and §Department of Molecular Biology, Princeton University, Princeton, NJ 08544

Contributed by David Botstein and Arnold J. Levine, October 21, 1998

**ABSTRACT** Wnt family members are critical to many developmental processes, and components of the Wnt signaling pathway have been linked to tumorigenesis in familial and sporadic colon carcinomas. Here we report the identification of two genes, *WISP-1* and *WISP-2*, that are up-regulated in the mouse mammary epithelial cell line C57MG transformed by Wnt-1, but not by Wnt-4. Together with a third related gene, *WISP-3*, these proteins define a subfamily of the connective tissue growth factor family. Two distinct systems demonstrated *WISP* induction to be associated with the expression of Wnt-1. These included (i) C57MG cells infected with a Wnt-1 retroviral vector or expressing Wnt-1 under the control of a tetracycline repressible promoter, and (ii) Wnt-1 transgenic mice. The *WISP-1* gene was localized to human chromosome 8q24.1-8q24.3. *WISP-1* genomic DNA was amplified in colon cancer cell lines and in human colon tumors and its RNA overexpressed (2- to >30-fold) in 84% of the tumors examined compared with patient-matched normal mucosa. *WISP-3* mapped to chromosome 6q22-6q23 and also was overexpressed (4- to >40-fold) in 63% of the colon tumors analyzed. In contrast, *WISP-2* mapped to human chromosome 20q12-20q13 and its DNA was amplified, but RNA expression was reduced (2- to >30-fold) in 79% of the tumors. These results suggest that the *WISP* genes may be downstream of Wnt-1 signaling and that aberrant levels of *WISP* expression in colon cancer may play a role in colon tumorigenesis.

Wnt-1 is a member of an expanding family of cysteine-rich, glycosylated signaling proteins that mediate diverse developmental processes such as the control of cell proliferation, adhesion, cell polarity, and the establishment of cell fates (1, 2). Wnt-1 originally was identified as an oncogene activated by the insertion of mouse mammary tumor virus in virus-induced mammary adenocarcinomas (3, 4). Although Wnt-1 is not expressed in the normal mammary gland, expression of Wnt-1 in transgenic mice causes mammary tumors (5).

In mammalian cells, Wnt family members initiate signaling by binding to the seven-transmembrane spanning Frizzled receptors and recruiting the cytoplasmic protein Dishevelled (Dsh) to the cell membrane (1, 2, 6). Dsh then inhibits the kinase activity of the normally constitutively active glycogen synthase kinase-3 $\beta$  (GSK-3 $\beta$ ) resulting in an increase in  $\beta$ -catenin levels. Stabilized  $\beta$ -catenin interacts with the transcription factor TCF/Lef1, forming a complex that appears in

the nucleus and binds TCF/Lef1 target DNA elements to activate transcription (7, 8). Other experiments suggest that the adenomatous polyposis coli (APC) tumor suppressor gene also plays an important role in Wnt signaling by regulating  $\beta$ -catenin levels (9). APC is phosphorylated by GSK-3 $\beta$ , binds to  $\beta$ -catenin, and facilitates its degradation. Mutations in either APC or  $\beta$ -catenin have been associated with colon carcinomas and melanomas, suggesting these mutations contribute to the development of these types of cancer, implicating the Wnt pathway in tumorigenesis (1).

Although much has been learned about the Wnt signaling pathway over the past several years, only a few of the transcriptionally activated downstream components activated by Wnt have been characterized. Those that have been described cannot account for all of the diverse functions attributed to Wnt signaling. Among the candidate Wnt target genes are those encoding the nodal-related 3 gene, *Xnr3*, a member of the transforming growth factor (TGF)- $\beta$  superfamily, and the homeobox genes, *engrailed*, *gooseoid*, *twist* (*Xtwn*), and *siamois* (2). A recent report also identifies *c-myc* as a target gene of the Wnt signaling pathway (10).

To identify additional downstream genes in the Wnt signaling pathway that are relevant to the transformed cell phenotype, we used a PCR-based cDNA subtraction strategy, suppression subtractive hybridization (SSH) (11), using RNA isolated from C57MG mouse mammary epithelial cells and C57MG cells stably transformed by a Wnt-1 retrovirus. Overexpression of Wnt-1 in this cell line is sufficient to induce a partially transformed phenotype, characterized by elongated and refractile cells that lose contact inhibition and form a multilayered array (12, 13). We reasoned that genes differentially expressed between these two cell lines might contribute to the transformed phenotype.

In this paper, we describe the cloning and characterization of two genes up-regulated in Wnt-1 transformed cells, *WISP-1* and *WISP-2*, and a third related gene, *WISP-3*. The *WISP* genes are members of the CCN family of growth factors, which includes connective tissue growth factor (CTGF), Cyr61, and *nov*, a family not previously linked to Wnt signaling.

### MATERIALS AND METHODS

**SSH.** SSH was performed by using the PCR-Select cDNA Subtraction Kit (CLONTECH). Tester double-stranded

Abbreviations: TGF, transforming growth factor; CTGF, connective tissue growth factor; SSH, suppression subtractive hybridization; VWC, von Willebrand factor type C module.

Data deposition: The sequences reported in this paper have been deposited in the Genbank database (accession nos. AF100777, AF100778, AF100779, AF100780, and AF100781).

†To whom reprint requests should be addressed. e-mail: diane@gene.com.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

© 1998 by The National Academy of Sciences 0027-8424/98/9514717-6\$2.00/0  
PNAS is available online at www.pnas.org.

cDNA was synthesized from 2  $\mu$ g of poly(A)<sup>+</sup> RNA isolated from the C57MG/Wnt-1 cell line and driver cDNA from 2  $\mu$ g of poly(A)<sup>+</sup> RNA from the parent C57MG cells. The subtracted cDNA library was subcloned into a pGEM-T vector for further analysis.

**cDNA Library Screening.** Clones encoding full-length mouse *WISP-1* were isolated by screening a  $\lambda$ gt10 mouse embryo cDNA library (CLONTECH) with a 70-bp probe from the original partial clone 568 sequence corresponding to amino acids 128–169. Clones encoding full-length human *WISP-1* were isolated by screening  $\lambda$ gt10 lung and fetal kidney cDNA libraries with the same probe at low stringency. Clones encoding full-length mouse and human *WISP-2* were isolated by screening a C57MG/Wnt-1 or human fetal lung cDNA library with a probe corresponding to nucleotides 1463–1512. Full-length cDNAs encoding *WISP-3* were cloned from human bone marrow and fetal kidney libraries.

**Expression of Human *WISP* RNA.** PCR amplification of first-strand cDNA was performed with human Multiple Tissue cDNA panels (CLONTECH) and 300  $\mu$ M of each dNTP at 94°C for 1 sec, 62°C for 30 sec, 72°C for 1 min, for 22–32 cycles. *WISP* and glyceraldehyde-3-phosphate dehydrogenase primer sequences are available on request.

**In Situ Hybridization.** <sup>33</sup>P-labeled sense and antisense riboprobes were transcribed from an 897-bp PCR product corresponding to nucleotides 601–1440 of mouse *WISP-1* or a 294-bp PCR product corresponding to nucleotides 82–375 of mouse *WISP-2*. All tissues were processed as described (40).

**Radiation Hybrid Mapping.** Genomic DNA from each hybrid in the Stanford G3 and Genebridge4 Radiation Hybrid Panels (Research Genetics, Huntsville, AL) and human and hamster control DNAs were PCR-amplified, and the results were submitted to the Stanford or Massachusetts Institute of Technology web servers.

**Cell Lines, Tumors, and Mucosa Specimens.** Tissue specimens were obtained from the Department of Pathology (University of Pittsburgh) for patients undergoing colon resection and from the University of Leeds, United Kingdom. Genomic DNA was isolated (Qiagen) from the pooled blood of 10 normal human donors, surgical specimens, and the following ATCC human cell lines: SW480, COLO 320DM, HT-29, WiDr, and SW403 (colon adenocarcinomas), SW620 (lymph node metastasis, colon adenocarcinoma), HCT 116 (colon carcinoma), SK-CO-1 (colon adenocarcinoma, ascites), and HM7 (a variant of ATCC colon adenocarcinoma cell line LS 174T). DNA concentration was determined by using Hoechst dye 33258 intercalation fluorimetry. Total RNA was prepared by homogenization in 7 M GuSCN followed by centrifugation over CsCl cushions or prepared by using RNeasy.

**Gene Amplification and RNA Expression Analysis.** Relative gene amplification and RNA expression of *WISPs* and *c-myc* in the cell lines, colorectal tumors, and normal mucosa were determined by quantitative PCR. Gene-specific primers and fluorogenic probes (sequences available on request) were designed and used to amplify and quantitate the genes. The relative gene copy number was derived by using the formula  $2^{(\Delta\Delta C_t)}$  where  $\Delta C_t$  represents the difference in amplification cycles required to detect the *WISP* genes in peripheral blood lymphocyte DNA compared with colon tumor DNA or colon tumor RNA compared with normal mucosal RNA. The  $\Delta$ -method was used for calculation of the SE of the gene copy number or RNA expression level. The *WISP*-specific signal was normalized to that of the glyceraldehyde-3-phosphate dehydrogenase housekeeping gene. All TaqMan assay reagents were obtained from Perkin-Elmer Applied Biosystems.

## RESULTS

**Isolation of *WISP-1* and *WISP-2* by SSH.** To identify Wnt-1-inducible genes, we used the technique of SSH using the

mouse mammary epithelial cell line C57MG and C57MG cells that stably express Wnt-1 (11). Candidate differentially expressed cDNAs (1,384 total) were sequenced. Thirty-nine percent of the sequences matched known genes or homologues, 32% matched expressed sequence tags, and 29% had no match. To confirm that the transcript was differentially expressed, semiquantitative reverse transcription-PCR and Northern analysis were performed by using mRNA from the C57MG and C57MG/Wnt-1 cells.

Two of the cDNAs, *WISP-1* and *WISP-2*, were differentially expressed, being induced in the C57MG/Wnt-1 cell line, but not in the parent C57MG cells or C57MG cells overexpressing Wnt-4 (Fig. 1A and B). Wnt-4, unlike Wnt-1, does not induce the morphological transformation of C57MG cells and has no effect on  $\beta$ -catenin levels (13, 14). Expression of *WISP-1* was up-regulated approximately 3-fold in the C57MG/Wnt-1 cell line and *WISP-2* by approximately 5-fold by both Northern analysis and reverse transcription-PCR.

An independent, but similar, system was used to examine *WISP* expression after Wnt-1 induction. C57MG cells expressing the *Wnt-1* gene under the control of a tetracycline-repressible promoter produce low amounts of Wnt-1 in the repressed state but show a strong induction of *Wnt-1* mRNA and protein within 24 hr after tetracycline removal (8). The levels of Wnt-1 and *WISP* RNA isolated from these cells at various times after tetracycline removal were assessed by quantitative PCR. Strong induction of Wnt-1 mRNA was seen as early as 10 hr after tetracycline removal. Induction of *WISP* mRNA (2- to 6-fold) was seen at 48 and 72 hr (data not shown). These data support our previous observations that show that *WISP* induction is correlated with Wnt-1 expression. Because the induction is slow, occurring after approximately 48 hr, the induction of *WISPs* may be an indirect response to Wnt-1 signaling.

cDNA clones of human *WISP-1* were isolated and the sequence compared with mouse *WISP-1*. The cDNA sequences of mouse and human *WISP-1* were 1,766 and 2,830 bp in length, respectively, and encode proteins of 367 aa, with predicted relative molecular masses of  $\approx 40,000$  ( $M_r$  40 K). Both have hydrophobic N-terminal signal sequences, 38 conserved cysteine residues, and four potential N-linked glycosylation sites and are 84% identical (Fig. 2A).

Full-length cDNA clones of mouse and human *WISP-2* were 1,734 and 1,293 bp in length, respectively, and encode proteins of 251 and 250 aa, respectively, with predicted relative molecular masses of  $\approx 27,000$  ( $M_r$  27 K) (Fig. 2B). Mouse and human *WISP-2* are 73% identical. Human *WISP-2* has no potential N-linked glycosylation sites, and mouse *WISP-2* has one at

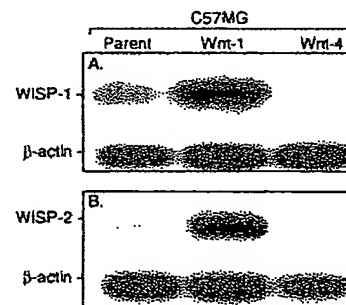


FIG. 1. *WISP-1* and *WISP-2* are induced by Wnt-1, but not Wnt-4, expression in C57MG cells. Northern analysis of *WISP-1* (A) and *WISP-2* (B) expression in C57MG, C57MG/Wnt-1, and C57MG/Wnt-4 cells. Poly(A)<sup>+</sup> RNA (2  $\mu$ g) was subjected to Northern blot analysis and hybridized with a 70-bp mouse *WISP-1*-specific probe (amino acids 278–300) or a 190-bp *WISP-2*-specific probe (nucleotides 1438–1627) in the 3' untranslated region. Blots were rehybridized with human  $\beta$ -actin probe.

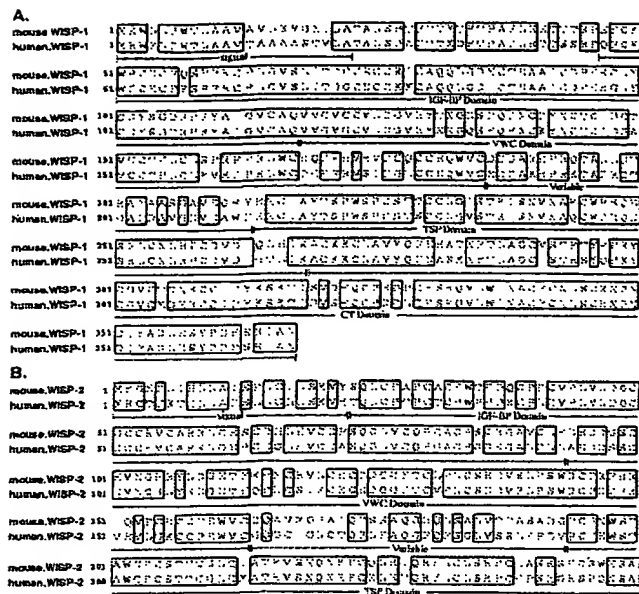


Fig. 2. Encoded amino acid sequence alignment of mouse and human *WISP-1* (A) and mouse and human *WISP-2* (B). The potential signal sequence, insulin-like growth factor-binding protein (IGF-BP), VWC, thrombospondin (TSP), and C-terminal (CT) domains are underlined.

position 197. *WISP-2* has 28 cysteine residues that are conserved among the 38 cysteines found in *WISP-1*.

**Identification of *WISP-3*.** To search for related proteins, we screened expressed sequence tag (EST) databases with the *WISP-1* protein sequence and identified several ESTs as potentially related sequences. We identified a homologous protein that we have called *WISP-3*. A full-length human *WISP-3* cDNA of 1,371 bp was isolated corresponding to those ESTs that encode a 354-aa protein with a predicted molecular mass of 39,293. *WISP-3* has two potential N-linked glycosylation sites and 36 cysteine residues. An alignment of the three human *WISP* proteins shows that *WISP-1* and *WISP-3* are the most similar (42% identity), whereas *WISP-2* has 37% identity with *WISP-1* and 32% identity with *WISP-3* (Fig. 3A).

***WISPs* Are Homologous to the CTGF Family of Proteins.** Human *WISP-1*, *WISP-2*, and *WISP-3* are novel sequences; however, mouse *WISP-1* is the same as the recently identified *Elm1* gene. *Elm1* is expressed in low, but not high, metastatic mouse melanoma cells, and suppresses the *in vivo* growth and metastatic potential of K-1735 mouse melanoma cells (15). Human and mouse *WISP-2* are homologous to the recently described rat gene, *rCop-1* (16). Significant homology (36–44%) was seen to the CCN family of growth factors. This family includes three members, CTGF, Cyr61, and the protooncogene *nov*. CTGF is a chemotactic and mitogenic factor for fibroblasts that is implicated in wound healing and fibrotic disorders and is induced by TGF- $\beta$  (17). Cyr61 is an extracellular matrix signaling molecule that promotes cell adhesion, proliferation, migration, angiogenesis, and tumor growth (18, 19). *nov* (nephroblastoma overexpressed) is an immediate early gene associated with quiescence and found altered in Wilms tumors (20). The proteins of the CCN family share functional, but not sequence, similarity to Wnt-1. All are secreted, cysteine-rich heparin binding glycoproteins that associate with the cell surface and extracellular matrix.

*WISP* proteins exhibit the modular architecture of the CCN family, characterized by four conserved cysteine-rich domains (Fig. 3B) (21). The N-terminal domain, which includes the first 12 cysteine residues, contains a consensus sequence (GCGC-CXXC) conserved in most insulin-like growth factor (IGF)-

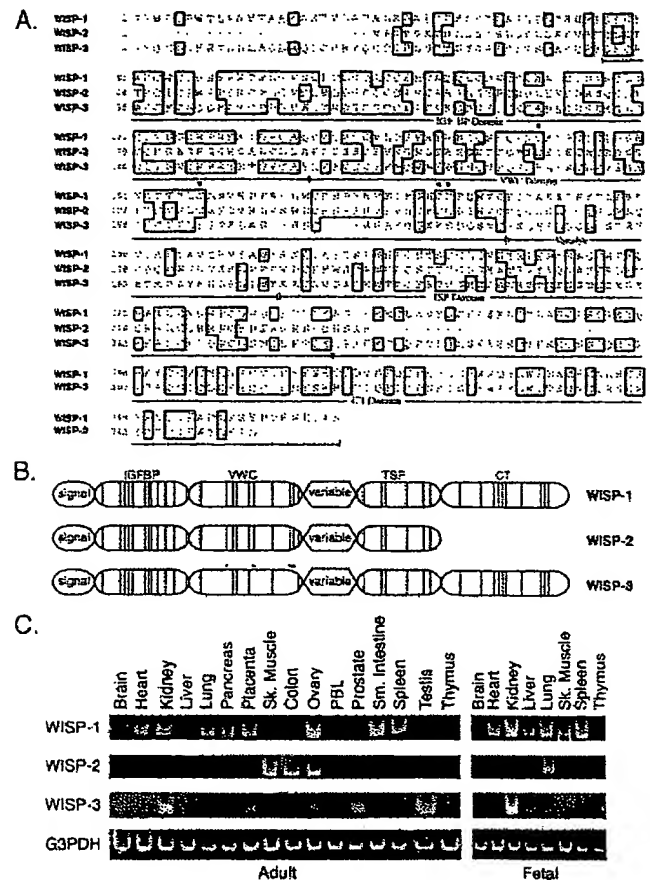


Fig. 3. (A) Encoded amino acid sequence alignment of human *WISPs*. The cysteine residues of *WISP-1* and *WISP-2* that are not present in *WISP-3* are indicated with a dot. (B) Schematic representation of the *WISP* proteins showing the domain structure and cysteine residues (vertical lines). The four cysteine residues in the VWC domain that are absent in *WISP-3* are indicated with a dot. (C) Expression of *WISP* mRNA in human tissues. PCR was performed on human multiple-tissue cDNA panels (CLONTECH) from the indicated adult and fetal tissues.

binding proteins (BP). This sequence is conserved in *WISP-2* and *WISP-3*, whereas *WISP-1* has a glutamine in the third position instead of a glycine. CTGF recently has been shown to specifically bind IGF (22) and a truncated *nov* protein lacking the IGF-BP domain is oncogenic (23). The von Willebrand factor type C module (VWC), also found in certain collagens and mucins, covers the next 10 cysteine residues, and is thought to participate in protein complex formation and oligomerization (24). The VWC domain of *WISP-3* differs from all CCN family members described previously, in that it contains only six of the 10 cysteine residues (Fig. 3A and B). A short variable region follows the VWC domain. The third module, the thrombospondin (TSP) domain is involved in binding to sulfated glycoconjugates and contains six cysteine residues and a conserved WSxCSxCG motif first identified in thrombospondin (25). The C-terminal (CT) module containing the remaining 10 cysteines is thought to be involved in dimerization and receptor binding (26). The CT domain is present in all CCN family members described to date but is absent in *WISP-2* (Fig. 3A and B). The existence of a putative signal sequence and the absence of a transmembrane domain suggest that *WISPs* are secreted proteins, an observation supported by an analysis of their expression and secretion from mammalian cell and baculovirus cultures (data not shown).

**Expression of *WISP* mRNA in Human Tissues.** Tissue-specific expression of human *WISPs* was characterized by PCR

analysis on adult and fetal multiple tissue cDNA panels. *WISP-1* expression was seen in the adult heart, kidney, lung, pancreas, placenta, ovary, small intestine, and spleen (Fig. 3C). Little or no expression was detected in the brain, liver, skeletal muscle, colon, peripheral blood leukocytes, prostate, testis, or thymus. *WISP-2* had a more restricted tissue expression and was detected in adult skeletal muscle, colon, ovary, and fetal lung. Predominant expression of *WISP-3* was seen in adult kidney and testis and fetal kidney. Lower levels of *WISP-3* expression were detected in placenta, ovary, prostate, and small intestine.

**In Situ Localization of *WISP-1* and *WISP-2*.** Expression of *WISP-1* and *WISP-2* was assessed by *in situ* hybridization in mammary tumors from Wnt-1 transgenic mice. Strong expression of *WISP-1* was observed in stromal fibroblasts lying within the fibrovascular tumor stroma (Fig. 4 A–D). However, low-level *WISP-1* expression also was observed focally within tumor cells (data not shown). No expression was observed in normal breast. Like *WISP-1*, *WISP-2* expression also was seen in the tumor stroma in breast tumors from Wnt-1 transgenic animals (Fig. 4 E–H). However, *WISP-2* expression in the stroma was in spindle-shaped cells adjacent to capillary vessels, whereas

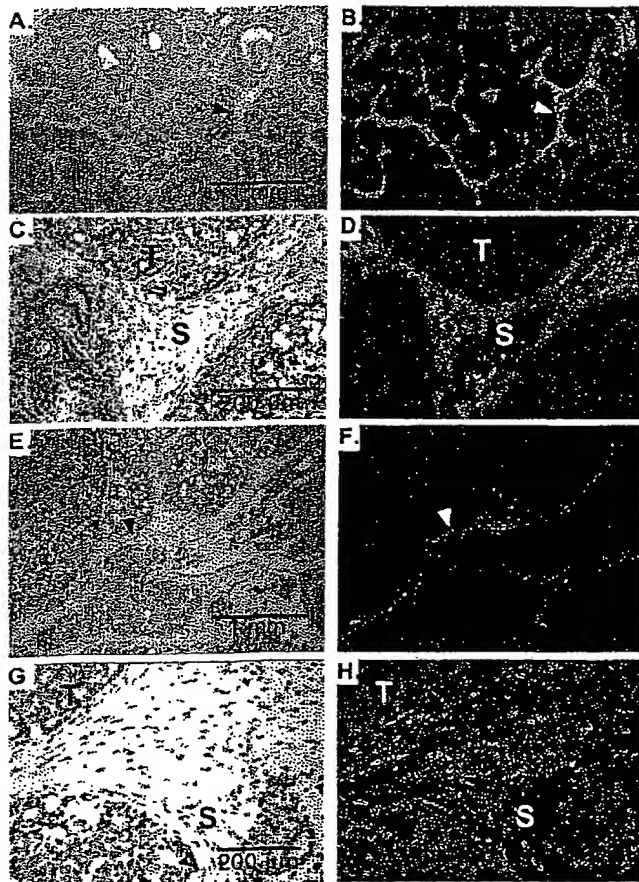


FIG. 4. (A, C, E, and G) Representative hematoxylin/eosin-stained images from breast tumors in Wnt-1 transgenic mice. The corresponding dark-field images showing *WISP-1* expression are shown in B and D. The tumor is a moderately well-differentiated adenocarcinoma showing evidence of adenoid cystic change. At low power (A and B), expression of *WISP-1* is seen in the delicate branching fibrovascular tumor stroma (arrowhead). At higher magnification, expression is seen in the stromal(s) fibroblasts (C and D), and tumor cells are negative. Focal expression of *WISP-1*, however, was observed in tumor cells in some areas. Images of *WISP-2* expression are shown in E–H. At low power (E and F), expression of *WISP-2* is seen in cells lying within the fibrovascular tumor stroma. At higher magnification, these cells appeared to be adjacent to capillary vessels whereas tumor cells are negative (G and H).

the predominant cell type expressing *WISP-1* was the stromal fibroblasts.

**Chromosomal Localization of the *WISP* Genes.** The chromosomal location of the human *WISP* genes was determined by radiation hybrid mapping panels. *WISP-1* is approximately 3.48 cR from the meiotic marker AFM259xc5 [logarithm of odds (lod) score 16.31] on chromosome 8q24.1 to 8q24.3, in the same region as the human locus of the *novH* family member (27) and roughly 4 Mbs distal to *c-myc* (28). Preliminary fine mapping indicates that *WISP-1* is located near D8S1712 STS. *WISP-2* is linked to the marker SHGC-33922 (lod = 1,000) on chromosome 20q12–20q13.1. Human *WISP-3* mapped to chromosome 6q22–6q23 and is linked to the marker AFM211ze5 (lod = 1,000). *WISP-3* is approximately 18 Mbs proximal to CTGF and 23 Mbs proximal to the human cellular oncogene *MYB* (27, 29).

**Amplification and Aberrant Expression of *WISPs* in Human Colon Tumors.** Amplification of protooncogenes is seen in many human tumors and has etiological and prognostic significance. For example, in a variety of tumor types, *c-myc* amplification has been associated with malignant progression and poor prognosis (30). Because *WISP-1* resides in the same general chromosomal location (8q24) as *c-myc*, we asked whether it was a target of gene amplification, and, if so, whether this amplification was independent of the *c-myc* locus. Genomic DNA from human colon cancer cell lines was assessed by quantitative PCR and Southern blot analysis. (Fig. 5 A and B). Both methods detected similar degrees of *WISP-1* amplification. Most cell lines showed significant (2- to 4-fold) amplification, with the HT-29 and WiDr cell lines demonstrating an 8-fold increase. Significantly, the pattern of amplification observed did not correlate with that observed for *c-myc*, indicating that the *c-myc* gene is not part of the amplicon that involves the *WISP-1* locus.

We next examined whether the *WISP* genes were amplified in a panel of 25 primary human colon adenocarcinomas. The relative *WISP* gene copy number in each colon tumor DNA was compared with pooled normal DNA from 10 donors by quantitative PCR (Fig. 6). The copy number of *WISP-1* and *WISP-2* was significantly greater than one, approximately 2-fold for *WISP-1* in about 60% of the tumors and 2- to 4-fold for *WISP-2* in 92% of the tumors ( $P < 0.001$  for each). The copy number for *WISP-3* was indistinguishable from one ( $P = 0.166$ ). In addition, the copy number of *WISP-2* was significantly higher than that of *WISP-1* ( $P < 0.001$ ).

The levels of *WISP* transcripts in RNA isolated from 19 adenocarcinomas and their matched normal mucosa were

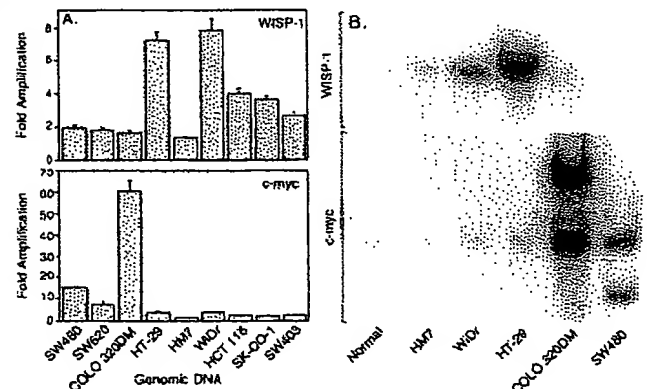


FIG. 5. Amplification of *WISP-1* genomic DNA in colon cancer cell lines. (A) Amplification in cell line DNA was determined by quantitative PCR. (B) Southern blots containing genomic DNA (10  $\mu$ g) digested with *EcoRI* (*WISP-1*) or *XbaI* (*c-myc*) were hybridized with a 100-bp human *WISP-1* probe (amino acids 186–219) or a human *c-myc* probe (located at bp 1901–2000). The *WISP* and *myc* genes are detected in normal human genomic DNA after a longer film exposure.

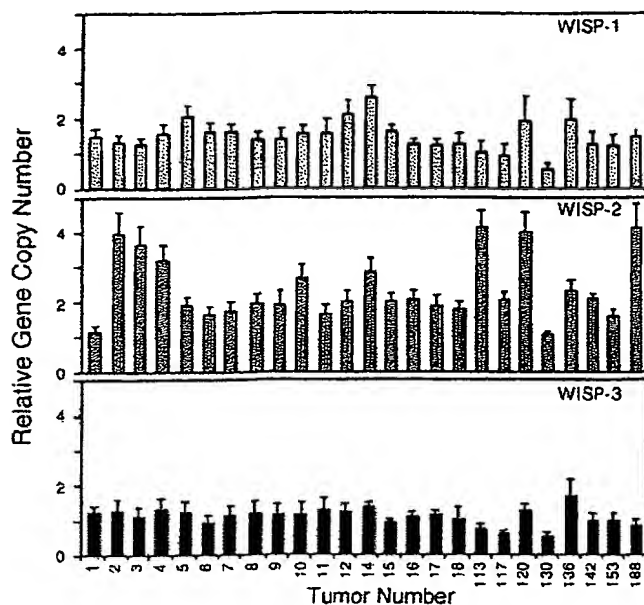


FIG. 6. Genomic amplification of *WISP* genes in human colon tumors. The relative gene copy number of the *WISP* genes in 25 adenocarcinomas was assayed by quantitative PCR, by comparing DNA from primary human tumors with pooled DNA from 10 healthy donors. The data are means  $\pm$  SEM from one experiment done in triplicate. The experiment was repeated at least three times.

assessed by quantitative PCR (Fig. 7). The level of *WISP-1* RNA present in tumor tissue varied but was significantly increased (2- to >25-fold) in 84% (16/19) of the human colon tumors examined compared with normal adjacent mucosa. Four of 19 tumors showed greater than 10-fold overexpression. In contrast, in 79% (15/19) of the tumors examined, *WISP-2* RNA expression was significantly lower in the tumor than the mucosa. Similar to *WISP-1*, *WISP-3* RNA was overexpressed in 63% (12/19) of the colon tumors compared with the normal

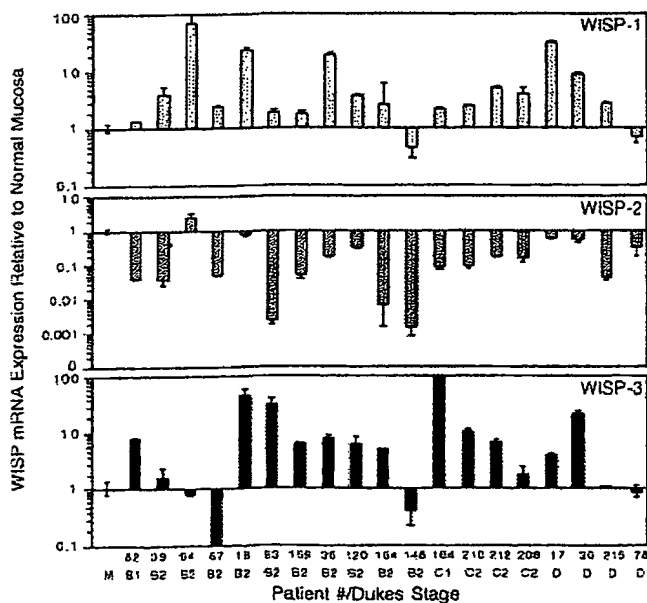


FIG. 7. *WISP* RNA expression in primary human colon tumors relative to expression in normal mucosa from the same patient. Expression of *WISP* mRNA in 19 adenocarcinomas was assayed by quantitative PCR. The Dukes stage of the tumor is listed under the sample number. The data are means  $\pm$  SEM from one experiment done in triplicate. The experiment was repeated at least twice.

mucosa. The amount of overexpression of *WISP-3* ranged from 4- to >40-fold.

## DISCUSSION

One approach to understanding the molecular basis of cancer is to identify differences in gene expression between cancer cells and normal cells. Strategies based on assumptions that steady-state mRNA levels will differ between normal and malignant cells have been used to clone differentially expressed genes (31). We have used a PCR-based selection strategy, SSH, to identify genes selectively expressed in C57MG mouse mammary epithelial cells transformed by Wnt-1.

Three of the genes isolated, *WISP-1*, *WISP-2*, and *WISP-3*, are members of the CCN family of growth factors, which includes CTGF, Cyr61, and *nov*, a family not previously linked to Wnt signaling.

Two independent experimental systems demonstrated that *WISP* induction was associated with the expression of Wnt-1. The first was C57MG cells infected with a Wnt-1 retroviral vector or C57MG cells expressing Wnt-1 under the control of a tetracycline-repressible promoter, and the second was in Wnt-1 transgenic mice, where breast tissue expresses Wnt-1, whereas normal breast tissue does not. No *WISP* RNA expression was detected in mammary tumors induced by polyoma virus middle T antigen (data not shown). These data suggest a link between Wnt-1 and *WISPs* in that in these two situations, *WISP* induction was correlated with Wnt-1 expression.

It is not clear whether the *WISPs* are directly or indirectly induced by the downstream components of the Wnt-1 signaling pathway (i.e.,  $\beta$ -catenin-TCF-1/Lef1). The increased levels of *WISP* RNA were measured in Wnt-1-transformed cells, hours or days after Wnt-1 transformation. Thus, *WISP* expression could result from Wnt-1 signaling directly through  $\beta$ -catenin transcription factor regulation or alternatively through Wnt-1 signaling turning on a transcription factor, which in turn regulates *WISPs*.

The *WISPs* define an additional subfamily of the CCN family of growth factors. One striking difference observed in the protein sequence of *WISP-2* is the absence of a CT domain, which is present in CTGF, Cyr61, *nov*, *WISP-1*, and *WISP-3*. This domain is thought to be involved in receptor binding and dimerization. Growth factors, such as TGF- $\beta$ , platelet-derived growth factor, and nerve growth factor, which contain a cystine knot motif exist as dimers (32). It is tempting to speculate that *WISP-1* and *WISP-3* may exist as dimers, whereas *WISP-2* exists as a monomer. If the CT domain is also important for receptor binding, *WISP-2* may bind its receptor through a different region of the molecule than the other CCN family members. No specific receptors have been identified for CTGF or *nov*. A recent report has shown that integrin  $\alpha_v\beta_3$  serves as an adhesion receptor for Cyr61 (33).

The strong expression of *WISP-1* and *WISP-2* in cells lying within the fibrovascular tumor stroma in breast tumors from Wnt-1 transgenic animals is consistent with previous observations that transcripts for the related CTGF gene are primarily expressed in the fibrous stroma of mammary tumors (34). Epithelial cells are thought to control the proliferation of connective tissue stroma in mammary tumors by a cascade of growth factor signals similar to that controlling connective tissue formation during wound repair. It has been proposed that mammary tumor cells or inflammatory cells at the tumor interstitial interface secrete TGF- $\beta$ 1, which is the stimulus for stromal proliferation (34). TGF- $\beta$ 1 is secreted by a large percentage of malignant breast tumors and may be one of the growth factors that stimulates the production of CTGF and *WISPs* in the stroma.

It was of interest that *WISP-1* and *WISP-2* expression was observed in the stromal cells that surrounded the tumor cells

(epithelial cells) in the Wnt-1 transgenic mouse sections of breast tissue. This finding suggests that paracrine signaling could occur in which the stromal cells could supply WISP-1 and WISP-2 to regulate tumor cell growth on the WISP extracellular matrix. Stromal cell-derived factors in the extracellular matrix have been postulated to play a role in tumor cell migration and proliferation (35). The localization of WISP-1 and WISP-2 in the stromal cells of breast tumors supports this paracrine model.

An analysis of WISP-1 gene amplification and expression in human colon tumors showed a correlation between DNA amplification and overexpression, whereas overexpression of WISP-3 RNA was seen in the absence of DNA amplification. In contrast, WISP-2 DNA was amplified in the colon tumors, but its mRNA expression was significantly reduced in the majority of tumors compared with the expression in normal colonic mucosa from the same patient. The gene for human WISP-2 was localized to chromosome 20q12–20q13, at a region frequently amplified and associated with poor prognosis in node negative breast cancer and many colon cancers, suggesting the existence of one or more oncogenes at this locus (36–38). Because the center of the 20q13 amplicon has not yet been identified, it is possible that the apparent amplification observed for WISP-2 may be caused by another gene in this amplicon.

A recent manuscript on *rCop-1*, the rat orthologue of WISP-2, describes the loss of expression of this gene after cell transformation, suggesting it may be a negative regulator of growth in cell lines (16). Although the mechanism by which WISP-2 RNA expression is down-regulated during malignant transformation is unknown, the reduced expression of WISP-2 in colon tumors and cell lines suggests that it may function as a tumor suppressor. These results show that the WISP genes are aberrantly expressed in colon cancer and suggest that their altered expression may confer selective growth advantage to the tumor.

Members of the Wnt signaling pathway have been implicated in the pathogenesis of colon cancer, breast cancer, and melanoma, including the tumor suppressor gene adenomatous polyposis coli and  $\beta$ -catenin (39). Mutations in specific regions of either gene can cause the stabilization and accumulation of cytoplasmic  $\beta$ -catenin, which presumably contributes to human carcinogenesis through the activation of target genes such as the WISPs. Although the mechanism by which Wnt-1 transforms cells and induces tumorigenesis is unknown, the identification of WISPs as genes that may be regulated downstream of Wnt-1 in C57MG cells suggests they could be important mediators of Wnt-1 transformation. The amplification and altered expression patterns of the WISPs in human colon tumors may indicate an important role for these genes in tumor development.

We thank the DNA synthesis group for oligonucleotide synthesis, T. Baker for technical assistance, P. Dowd for radiation hybrid mapping, K. Willert and R. Nusse for the tet-repressible C57MG/Wnt-1 cells, V. Dixit for discussions, and D. Wood and A. Bruce for artwork.

- Cadigan, K. M. & Nusse, R. (1997) *Genes Dev.* 11, 3286–3305.
- Dale, T. C. (1998) *Biochem. J.* 329, 209–223.
- Nusse, R. & Varmus, H. E. (1982) *Cell* 31, 99–109.
- van Ooyen, A. & Nusse, R. (1984) *Cell* 39, 233–240.
- Tsukamoto, A. S., Grosschedl, R., Guzman, R. C., Parslow, T. & Varmus, H. E. (1988) *Cell* 55, 619–625.
- Brown, J. D. & Moon, R. T. (1998) *Curr. Opin. Cell Biol.* 10, 182–187.
- Molenaar, M., van de Wetering, M., Oosterwegel, M., Peterson-Maduro, J., Godsave, S., Korinek, V., Roose, J., Destree, O. & Clevers, H. (1996) *Cell* 86, 391–399.
- Korinek, V., Barker, N., Willert, K., Molenaar, M., Roose, J., Wagenaar, G., Markman, M., Lamers, W., Destree, O. & Clevers, H. (1998) *Mol. Cell Biol.* 18, 1248–1256.
- Munemitsu, S., Albert, I., Souza, B., Rubinfeld, B. & Polakis, P. (1995) *Proc. Natl. Acad. Sci. USA* 92, 3046–3050.
- He, T. C., Sparks, A. B., Rago, C., Hermeking, H., Zawel, L., da Costa, L. T., Morin, P. J., Vogelstein, B. & Kinzler, K. W. (1998) *Science* 281, 1509–1512.
- Diatchenko, L., Lau, Y. F., Campbell, A. P., Chenchik, A., Moqadam, F., Huang, B., Lukyanov, S., Lukyanov, K., Gurskaya, N., Sverdlov, E. D. & Siebert, P. D. (1996) *Proc. Natl. Acad. Sci. USA* 93, 6025–6030.
- Brown, A. M., Wildin, R. S., Prendergast, T. J. & Varmus, H. E. (1986) *Cell* 46, 1001–1009.
- Wong, G. T., Gavin, B. J. & McMahon, A. P. (1994) *Mol. Cell Biol.* 14, 6278–6286.
- Shimizu, H., Julius, M. A., Giarre, M., Zheng, Z., Brown, A. M. & Kitajewski, J. (1997) *Cell Growth Differ.* 8, 1349–1358.
- Hashimoto, Y., Shindo-Okada, N., Tani, M., Nagamachi, Y., Takeuchi, K., Shiroishi, T., Toma, H. & Yokota, J. (1998) *J. Exp. Med.* 187, 289–296.
- Zhang, R., Averboukh, L., Zhu, W., Zhang, H., Jo, H., Dempsey, P. J., Coffey, R. J., Pardee, A. B. & Liang, P. (1998) *Mol. Cell Biol.* 18, 6131–6141.
- Grotendorst, G. R. (1997) *Cytokine Growth Factor Rev.* 8, 171–179.
- Kireeva, M. L., Mo, F. E., Yang, G. P. & Lau, L. F. (1996) *Mol. Cell Biol.* 16, 1326–1334.
- Babic, A. M., Kireeva, M. L., Kolesnikova, T. V. & Lau, L. F. (1998) *Proc. Natl. Acad. Sci. USA* 95, 6355–6360.
- Martinerie, C., Huff, V., Joubert, I., Badzioch, M., Saunders, G., Strong, L. & Perbal, B. (1994) *Oncogene* 9, 2729–2732.
- Bork, P. (1993) *FEBS Lett.* 327, 125–130.
- Kim, H. S., Nagalla, S. R., Oh, Y., Wilson, E., Roberts, C. T., Jr. & Rosenfeld, R. G. (1997) *Proc. Natl. Acad. Sci. USA* 94, 12981–12986.
- Joliet, V., Martinerie, C., Dambrine, G., Plassiart, G., Brisac, M., Crochet, J. & Perbal, B. (1992) *Mol. Cell Biol.* 12, 10–21.
- Mancuso, D. J., Tuley, E. A., Westfield, L. A., Worrall, N. K., Shelton-Inloes, B. B., Sorace, J. M., Alevis, Y. G. & Sadler, J. E. (1989) *J. Biol. Chem.* 264, 19514–19527.
- Holt, G. D., Pangburn, M. K. & Ginsburg, V. (1990) *J. Biol. Chem.* 265, 2852–2855.
- Voorberg, J., Fontijn, R., Calafat, J., Janssen, H., van Mourik, J. A. & Pannekoek, H. (1991) *J. Cell Biol.* 113, 195–205.
- Martinerie, C., Viegas-Pequignot, E., Guenard, I., Dutrillaux, B., Nguyen, V. C., Bernheim, A. & Perbal, B. (1992) *Oncogene* 7, 2529–2534.
- Takahashi, E., Hori, T., O'Connell, P., Leppert, M. & White, R. (1991) *Cytogenet. Cell Genet.* 57, 109–111.
- Meese, E., Meltzer, P. S., Witkowski, C. M. & Trent, J. M. (1989) *Genes Chromosomes Cancer* 1, 88–94.
- Garte, S. J. (1993) *Crit. Rev. Oncog.* 4, 435–449.
- Zhang, L., Zhou, W., Velculescu, V. E., Kern, S. E., Hruban, R. H., Hamilton, S. R., Vogelstein, B. & Kinzler, K. W. (1997) *Science* 276, 1268–1272.
- Sun, P. D. & Davies, D. R. (1995) *Annu. Rev. Biophys. Biomol. Struct.* 24, 269–291.
- Kireeva, M. L., Lam, S. C. T. & Lau, L. F. (1998) *J. Biol. Chem.* 273, 3090–3096.
- Frazier, K. S. & Grotendorst, G. R. (1997) *Int. J. Biochem. Cell Biol.* 29, 153–161.
- Wernert, N. (1997) *Virchows Arch.* 430, 433–443.
- Tanner, M. M., Tirkkonen, M., Kallioniemi, A., Collins, C., Stokke, T., Karhu, R., Kowbel, D., Shadravan, F., Hintz, M., Kuo, W. L., et al. (1994) *Cancer Res.* 54, 4257–4260.
- Brinkmann, U., Gallo, M., Polymeropoulos, M. H. & Pastan, I. (1996) *Genome Res.* 6, 187–194.
- Bischoff, J. R., Anderson, L., Zhu, Y., Mossie, K., Ng, L., Souza, B., Schryver, B., Flanagan, P., Clairvoyant, F., Ginther, C., et al. (1998) *EMBO J.* 17, 3052–3065.
- Morin, P. J., Sparks, A. B., Korinek, V., Barker, N., Clevers, H., Vogelstein, B. & Kinzler, K. W. (1997) *Science* 275, 1787–1790.
- Lu, L. H. & Gillett, N. (1994) *Cell Vision* 1, 169–176.

## Overexpression of a DEAD Box Protein (DDX1) in Neuroblastoma and Retinoblastoma Cell Lines\*

(Received for publication, November 17, 1997, and in revised form, June 2, 1998)

Roseline Godbout†, Mary Packer, and Wenjun Bie

From the Department of Oncology, Cross Cancer Institute and University of Alberta, 11560 University Ave., Edmonton, Alberta T6G1Z2, Canada

The DEAD box gene, *DDX1*, is a putative RNA helicase that is co-amplified with *MYCN* in a subset of retinoblastoma (RB) and neuroblastoma (NB) tumors and cell lines. Although gene amplification usually involves hundreds to thousands of kilobase pairs of DNA, a number of studies suggest that co-amplified genes are only overexpressed if they provide a selective advantage to the cells in which they are amplified. Here, we further characterize *DDX1* by identifying its putative transcription and translation initiation sites. We analyze *DDX1* protein levels in *MYCN/DDX1*-amplified NB and RB cell lines using polyclonal antibodies specific to *DDX1* and show that there is a good correlation with *DDX1* gene copy number, *DDX1* transcript levels, and *DDX1* protein levels in all cell lines studied. *DDX1* protein is found in both the nucleus and cytoplasm of *DDX1*-amplified lines but is localized primarily to the nucleus of nonamplified cells. Our results indicate that *DDX1* may be involved in either the formation or progression of a subset of NB and RB tumors and suggest that *DDX1* normally plays a role in the metabolism of RNAs located in the nucleus of the cell.

DEAD box proteins are a family of putative RNA helicases that are characterized by eight conserved amino acid motifs, one of which is the ATP hydrolysis motif containing the core amino acid sequence DEAD (Asp-Glu-Ala-Asp) (1–3). Over 40 members of the DEAD box family have been isolated from a variety of organisms including bacteria, yeast, insects, amphibians, mammals, and plants. The prototypic DEAD box protein is the translation initiation factor, eukaryotic initiation factor 4A, which, when combined with eukaryotic initiation factor 4B, unwinds double-stranded RNA (4). Other DEAD box proteins, such as p68, Vasa, and An3, can effectively and independently destabilize/unwind short RNA duplexes *in vitro* (5–7). Although some DEAD box proteins play general roles in cellular processes such as translation initiation (eukaryotic initiation factor 4A (4)), RNA splicing (PRP5, PRP28, and SPP81 in yeast (8–10)), and ribosomal assembly (SrmB in *Escherichia coli* (11)), the function of most DEAD box proteins remains unknown. Many of the DEAD box proteins found in higher eukaryotes are tissue- or stage-specific. For example, *PL10* mRNA is expressed only in the male germ line, and its product

has been proposed to have a specific role in translational regulation during spermatogenesis (12). Vasa and ME31B are maternal proteins that may be involved in embryogenesis (13, 14). p68, found in dividing cells (15), is believed to be required for the formation of nucleoli and may also have a function in the regulation of cell growth and division (16, 17). Other DEAD box proteins are implicated in RNA degradation, mRNA stability, and RNA editing (18–20).

The human DEAD box protein gene *DDX1*<sup>1</sup> was identified by differential screening of a cDNA library enriched in transcripts present in the two RB cell lines Y79 and RB522A (21). The longest *DDX1* cDNA insert isolated from this library was 2.4 kb with an open reading frame from position 1 to 2201. All eight conserved motifs characteristic of DEAD box proteins are found in the predicted amino acid sequence of *DDX1* as well as a region with homology to the heterogeneous nuclear ribonucleoprotein U, a protein believed to participate in the processing of heterogeneous nuclear RNA to mRNA (22, 23). The region of homology to heterogeneous nuclear ribonucleoprotein U spans 128 amino acids and is located between the first two conserved DEAD box protein motifs, 1a and 1b.

The proto-oncogene *MYCN* encodes a member of the MYC family of transcription factors that bind to an E box element (CACGTG) when dimerized with the MAX protein (24, 25). The *MYCN* gene is amplified and overexpressed in approximately one-third of all NB tumors (26, 27). Amplification of *MYCN* is associated with rapid tumor progression and a poor clinical prognosis (26, 27). *MYCN* overexpression is usually achieved by increasing gene copy number rather than by up-regulating basal expression of *MYCN* (27, 28). Because gene amplification involves hundreds to thousands of kilobase pairs of contiguous DNA (29–32), it is possible that co-amplification of a gene located in proximity to *MYCN* may contribute to the poor clinical prognosis of *MYCN*-amplified tumors. The *DDX1* gene maps to the same chromosomal band as *MYCN*, 2p24, and is located ~400 kb telomeric to the *MYCN* gene (33–36). All four *MYCN*-amplified RB tumor cell lines tested to date are amplified for *DDX1* (21),<sup>2</sup> while approximately two-thirds of NB cell lines and 38–68% of NB tumors are co-amplified for both genes (37–39). George *et al.* (39) found a significant decrease in the mean disease-free survival of patients with *DDX1*/*MYCN*-amplified NB tumors compared with *MYCN*-amplified tumors. Similarly, Squire *et al.* (38) observed a trend toward a worse clinical prognosis when both genes were amplified in the tumors of NB patients. To date, there have been no reports of a

\* This work was supported by the National Cancer Institute of Canada with funds from the Canadian Cancer Society. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) X70649.

† To whom correspondence should be addressed: Dept. of Oncology, Cross Cancer Institute, 11560 University Ave., Edmonton, Alberta T6G 1Z2, Canada. Tel.: 403-432-8901; Fax: 403-432-8892.

<sup>1</sup> The abbreviations used are: *DDX1*, DEAD box 1; NB, neuroblastoma; RB, retinoblastoma; RACE, rapid amplification of cDNA ends; PAGE, polyacrylamide gel electrophoresis; nt, nucleotide(s); MOPS, 4-morpholinepropanesulfonic acid; bp, base pair(s); kb, kilobase(s) or kilobase pair(s).

<sup>2</sup> R. Godbout, unpublished results.

tumor amplified only for *DDX1*, and the role that this gene plays in cancer formation and progression is not known.

Because of the high rate of rearrangements in amplified DNA (31, 40), it is unlikely that a gene located ~400 kb from the *MYCN* gene will be consistently amplified as an intact unit unless its product provides a growth advantage to the cell. Based on Southern blot analysis, the *DDX1* gene extends over more than 30 kb, and there are no gross rearrangements of this gene in *DDX1*-amplified tumors (21, 38). Furthermore, there is a good correlation between *DDX1* transcript levels and gene copy number in the tumors analyzed to date. However, we need to show that *DDX1* protein is overexpressed in *DDX1*-amplified tumors if we are to entertain the possibility that this protein plays a role in the tumorigenic process. Here, we isolate and characterize the 5'-end of *DDX1* mRNA and extend the *DDX1* cDNA sequence by ~300 nt. We identify the predicted initiation codon of *DDX1* and generate antisera that specifically recognize *DDX1* protein. We analyze levels of *DDX1* protein in both *DDX1*-amplified and nonamplified RB and NB tumors and study the subcellular location of this protein in the cell.

#### MATERIALS AND METHODS

**Library Screening**—A human fetal brain cDNA library (Stratagene) was screened using a 320-bp DNA fragment from the 5'-end of the 2.4-kb *DDX1* cDNA previously described (23). Phagemids containing positive inserts were excised from  $\lambda$  ZAP II following the supplier's directions. The ends of the cDNA inserts were sequenced using the dideoxynucleotide chain termination method with T7 DNA polymerase (Amersham Pharmacia Biotech).

A human placenta genomic library (CLONTECH) was screened with the 5'-end of *DDX1* cDNA. Positive plaques were purified, and the genomic DNA was analyzed using restriction enzymes and Southern blotting. *EcoRI*-digested DNA fragments from these clones were subcloned into pBluescript and digested with exonuclease III and mung bean nuclease to obtain sequentially deleted clones. The exon/intron map of the 5' portion of the *DDX1* gene was obtained by comparing the sequence of *DDX1* cDNA with that of the genomic DNA.

**Rapid Amplification of cDNA Ends (RACE)**—We used the AmpliFINDER RACE kit (CLONTECH) to extend the 5'-end of *DDX1* cDNA. Briefly, two  $\mu$ g of poly(A)<sup>+</sup> RNA isolated from RB522A was reverse transcribed at 52 °C using either primer P1 or P3 (Fig. 1A). The RNA template was hydrolyzed, and excess primer was removed. A single-stranded AmpliFINDER anchor containing an *EcoRI* site was ligated to the 3'-end of the cDNA using T4 RNA ligase. The cDNA was amplified using either primer P2 or P4 (Fig. 1A) and AmpliFINDER anchor primer. RACE products were cloned into pBluescript.

**Primer Extension**—Poly(A)<sup>+</sup> RNAs were isolated from RB and NB cell lines as described previously (21, 38). The 21-nt primers 5'-TTTCGTCTGGGCACCATGTGT-3' (primer P4 in Fig. 1A) and 5'-TGGGACCTAGGGCTTCTGGAC-3' (primer P3 in Fig. 1A) were end-labeled with [ $\gamma$ -<sup>32</sup>P]ATP (3000 Ci/mmol; Mandel Scientific) and T4 polynucleotide kinase. Each of the labeled primers was annealed to 2  $\mu$ g of poly(A)<sup>+</sup> RNA at 45 °C for 90 min, and the cDNA was extended at 42 °C for 60 min using avian myeloblastosis virus reverse transcriptase (Promega). The primer extension products were heat-denatured and run on a 8% polyacrylamide gel containing 7 M urea in 1 $\times$  TBE buffer. A G + A sequencing ladder served as the size standard.

**S1 Nuclease Protection Assay**—The S1 nuclease protection assay to map the transcription initiation site of *DDX1* was performed as described by Favaloro *et al.* (41). The DNA probe was prepared by digesting genomic DNA spanning the upstream region of *DDX1* and exon 1 with *AvaI*, labeling the ends with [ $\gamma$ -<sup>32</sup>P]ATP (3000 Ci/mmol) and polynucleotide kinase, and removing the label from one of the ends by digesting the DNA with *SphI* (Fig. 4). The RNA samples were resuspended in a hybridization mixture containing 80% formamide, 40 mM PIPES, 400 mM NaCl, 1 mM EDTA, and the heat-denatured *SphI*-*AvaI* probe labeled at the *AvaI* site. The samples were incubated at 45 °C for 16 h and digested with 3000 units/ml S1 nuclease (Boehringer Mannheim) for 60 min at 37 °C. The samples were precipitated with ethanol; resuspended in 80% formaldehyde, TBE buffer, 0.1% bromophenol blue, xylene cyanol; denatured at 90 °C for 2 min; and electrophoresed in a 7 M urea, 8% polyacrylamide gel in TBE buffer.

**Northern and Southern Blot Analysis**—Poly(A)<sup>+</sup> RNAs were isolated from RB and NB cell lines as described previously (21, 38). Two  $\mu$ g of

poly(A)<sup>+</sup> RNA/lane were electrophoresed in a 6% formaldehyde, 1.5% agarose gel in MOPS buffer (20 mM MOPS, 5 mM sodium acetate, 1 mM EDTA, pH 7.0) and transferred to nitrocellulose filter in 3 M sodium chloride, 0.3 M sodium citrate. The filters were hybridized to the following DNA probes, <sup>32</sup>P-labeled by nick translation: (i) a 1.6-kb *EcoRI* insert from *DDX1* cDNA clone 1042 (21), (ii) a 260-bp cDNA fragment spanning the 3'-end of *DDX1* exon 1 as well as exons 2 and 3, (iii) a 160-bp fragment derived from the 5'-end of *DDX1* exon 1, and (iv)  $\alpha$ -actin cDNA to control for lane to lane variation in RNA levels. Filters were hybridized and washed under high stringency. Southern blot analysis was as described previously (21).

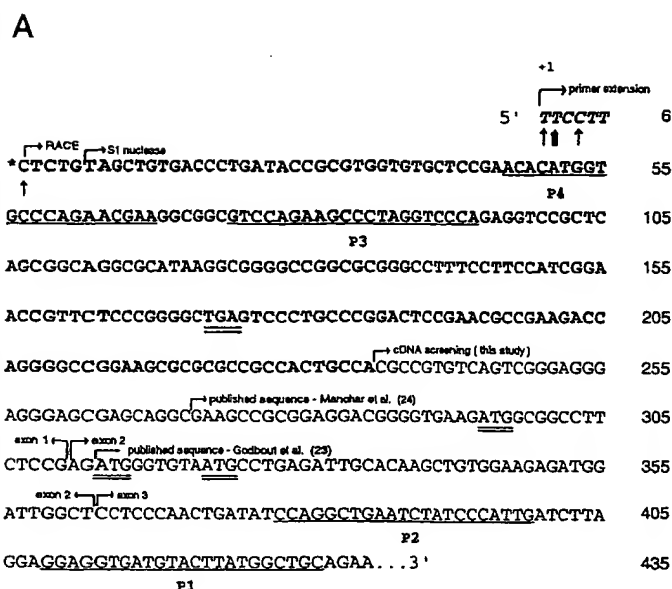
**Preparation of Anti-DDX1 Antiserum**—To prepare antiserum to the C terminus of the *DDX1* protein, we inserted a 1.8-kb *EcoRI* fragment from bp 848 to 2668 of *DDX1* cDNA (Fig. 1B) into *EcoRI*-digested pMAL-c2 expression vector (New England Biolabs). DH5 $\alpha$  cells transformed with this vector were grown to mid-log phase and induced with 0.1 mM isopropyl-1-thio- $\beta$ -D-thiogalactoside. The cells were harvested 3–4 h postinduction and lysed by sonication. Soluble maltose binding protein-DDX1 fusion protein was affinity-purified using amylose resin, and the maltose-binding protein was cleaved with factor Xa. The *DDX1* protein was purified on a SDS-PAGE gel, electroeluted, and concentrated. Approximately 100  $\mu$ g of protein was injected into rabbits at 4–6-week intervals. For the initial injection, the protein was dispersed in complete Freund's adjuvant (Sigma), while subsequent injections were prepared in Freund's incomplete adjuvant. Blood was collected from each rabbit 10 days after injection, and the specificity of the antiserum was tested using cell extracts from RB522A. To prepare antiserum to the N terminus of *DDX1* protein, a *DDX1* cDNA fragment from bp 268 to 851 (Fig. 1B) was inserted into pGEX-4T2 (Amersham Pharmacia Biotech). The recombinant protein produced from this construct contains the first 186 amino acids of the predicted *DDX1* sequence. Soluble glutathione S-transferase-DDX1 fusion protein was purified with glutathione-Sepharose 4B (Amersham Pharmacia Biotech). The glutathione S-transferase component of the fusion protein was cleaved with thrombin.

**Subcellular Fractionations and Western Blot Analysis**—We used two different procedures for subcellular fractionations. First, we isolated nuclear and S100 (soluble cytoplasmic) fractions from RB522A, IMR-32, Y79, RB(E)-2, HeLa, and HL60 using the procedure of Dignam (42). On average, we obtained 5–6 times more protein in the cytosolic fractions than in the nuclear fractions. Second, 10<sup>8</sup> RB522A cells were lysed and fractionated into S4 (soluble cytoplasmic components), P2 (heavy mitochondria, plasma membrane fragments), P3 (mitochondria, lysosomes, peroxisomes, and Golgi membranes), and P4 fractions (membrane vesicles from rough and smooth endoplasmic reticulum, Golgi, and plasma membrane) by differential centrifugation (43). We obtained 8 mg of protein in the S4 fraction, 1 mg in P2, 0.5 mg in P3, and 2 mg in P4 fraction. The procedures related to the immunoelectron microscopy have been previously described (44).

For Western blot analysis, proteins were electrophoresed in polyacrylamide-SDS gels and electroblotted onto nitrocellulose using the standard protocol for protein transfer described by Schleicher and Schuell. The filters were incubated with a 1:5000 dilution of *DDX1* antiserum, a 1:200 dilution of anti-MYCYN monoclonal antibody (Boehringer Mannheim), or a 1:200 dilution of anti-actin (Santa Cruz Biotechnology, Inc., Santa Cruz, CA). For the colorimetric analysis, antigen-antibody interactions were visualized using either alkaline phosphatase-linked goat anti-rabbit IgG (for *DDX1*) or goat anti-mouse IgG (for MYCYN) at a 1:3000 dilution. For the ECL Western blotting analysis (Amersham Pharmacia Biotech), we used a 1:100,000 dilution of peroxidase-linked secondary anti-rabbit IgG antibody (for *DDX1*) or secondary anti-goat IgG antibody (Jackson ImmunoResearch Laboratories).

#### RESULTS

**Identification of the 5'-End of the *DDX1* Transcript**—We have previously reported the sequence of *DDX1* cDNA isolated from an RB cDNA library (21, 23). This 2.4-kb *DDX1* cDNA contains an open reading frame spanning positions 1–2201 with a methionine encoded by the first three nucleotides (Fig. 1A). There is a polyadenylation signal and poly(A) tail in the 3'-untranslated region, indicating that the sequence is complete at the 3'-end. Manohar *et al.* (37) have also isolated *DDX1* cDNA from the NB cell line LA-N-5. Their cDNA extended the 5'-end of our sequence by 42 bp and included an additional in frame methionine (*double underlined* in Fig. 1A). The possibil-



**FIG. 1. Partial sequence and structure of *DDX1* cDNA.** A, the sequence of the 5'-end of *DDX1* cDNA. The sequence in **boldface type** starting at the *asterisk* was obtained using the RACE strategy. The additional 6 bp in **italic boldface type** at the 5'-end of the cDNA are predicted based on the known *DDX1* genomic sequence and primer extension analysis. P1, P2, P3, and P4 are primers used in the RACE experiments (the complementary sequence was used in each case). Primers P3 and P4 were also used for the primer extension analysis. Three in frame methionine codons are indicated by the **double underline**. An in frame stop codon is indicated by the **boldface double underline**. The three major transcription initiation sites identified by primer extension are indicated by the **single arrows**, while a minor site is represented by the **broad arrow**. The predicted *DDX1* transcription initiation sites obtained by RACE, S1 nuclease, and primer extension are indicated as well as the 5'-ends of *DDX1* cDNA sequences obtained by screening cDNA libraries. The sequences transcribed from exons 1, 2, and 3 are also shown. B, the structure of the 2711-bp *DDX1* cDNA is shown with an open reading frame from position 295 to 2515.

ity of additional in frame methionines located further upstream could not be excluded, because there were no predicted stop codons in the upstream region of the cDNA.

Northern blot analysis indicated a *DDX1* transcript size of ~2800 nt, suggesting that the *DDX1* cDNAs isolated to date were lacking ~300–350 bp of 5' sequence. We have used different approaches to identify the transcription start site of *DDX1*. First, we exhaustively screened a commercial fetal brain cDNA library with the 5'-end of *DDX1* cDNA. Although numerous clones were analyzed, only one extended the sequence (by 35 bp) beyond that published by Manohar *et al.* (37) (Fig. 1A).

We next used the RACE procedure in an attempt to isolate additional 5' sequence. The nested primers used to amplify the 5'-end of the *DDX1* transcript are labeled as primers P1 and P2 in Fig. 1A and are located downstream of the three in frame methionines (**double underline** in Fig. 1A). Poly(A)<sup>+</sup> RNA from RB522A was reverse transcribed at 52 °C using primer P1, and the reverse transcribed cDNA was amplified using the nested primer P2 and the 5'-RACE primer. Using this approach, we generated a product that was 230 bp longer than any of the cDNAs obtained by screening libraries (Fig. 1A). Sequencing of this 230-bp cDNA revealed an in frame stop codon (**boldface double underline** in Fig. 1A) located 123 bp



**FIG. 2. Identification of the 5'-end of the *DDX1* transcript by primer extension.** Radioactively labeled primer P4 was annealed to 2 µg of poly(A)<sup>+</sup> RNA from RB522A (lane 1), 1 µg of poly(A)<sup>+</sup> RNA from RB522A (lane 2), and 2 µg of poly(A)<sup>+</sup> RNA from RB(E)-2 cells (lane 3), and extended using reverse transcriptase. The products were run on an 8% denaturing polyacrylamide gel with a G + A sequencing ladder as size marker. The primer extension products are indicated on the left. The sizes of the products (in nt) are presented as the distance from primer P4.

upstream of the predicted translation initiation site. We then prepared primers P3 and P4, located near the 5'-end of the RACE cDNA (Fig. 1A) and repeated the RACE procedure to see if additional 5' sequences could be obtained. The resulting RACE products did not extend the *DDX1* cDNA sequence further.

The location of the *DDX1* transcription initiation site was verified by primer extension. Poly(A)<sup>+</sup> RNA was prepared from the following two cell lines: *DDX1*-amplified RB cell line RB522A and a nonamplified RB cell line RB(E)-2. RB522A has elevated levels of *DDX1* mRNA, while RB(E)-2 has at least 20-fold lower levels of this transcript. Three products of 40, 43, and 46 nt (with a weak signal at 45 nt) were detected in RB522A using primer P4 (Figs. 1A and 2). The 40-nt product corresponded exactly with the 5'-end of the RACE-derived cDNA while the 43- and 46-nt products extended the predicted size of the *DDX1* transcript by 3 and 6 nt, respectively. None of these products were observed in RB(E)-2. Bands of identical sizes to those obtained with RB522A mRNA were also observed in the *DDX1*-amplified NB cell line BE(2)-C but not in the *DDX1*-amplified NB cell line IMR-32 (data not shown). The same predicted *DDX1* transcription initiation site was identified with primer P3 except that the bands were of weaker

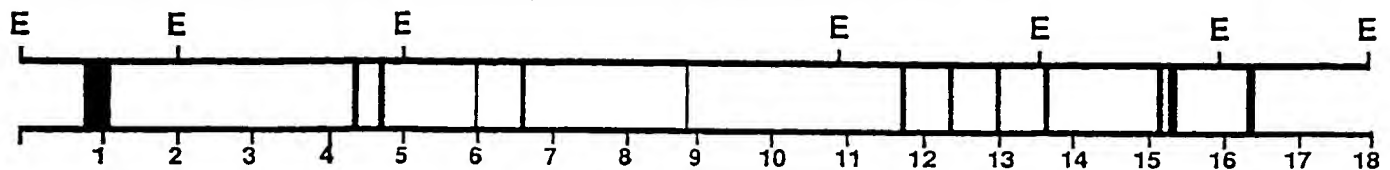


FIG. 3. Genomic map of the 5'-end of *DDX1*. The exons are represented by the black boxes, and distances are in kilobase pairs. The locations of *EcoRI* (E) sites are indicated.

intensity (data not shown). We have designated the transcription start site identified by primer extension as +1 (Fig. 1A).

The sequence of the 6 nt extending beyond the RACE cDNA was obtained by comparison of the cDNA sequence with that of *DDX1* genomic DNA. Bacteriophages containing *DDX1* genomic DNA were isolated by screening a human placenta library with 5' *DDX1* cDNA. Eighteen kb of DNA were sequenced from two bacteriophages with overlapping *DDX1* genomic DNA. Thirteen exons were identified within this 18-kb region (Fig. 3) corresponding to cDNA sequences from position 1 to 1249. The 310-bp exon 1 was by far the longest of the 13 exons sequenced, corresponding to the entire 5'-untranslated region of *DDX1* as well as the first in frame methionine. The sequences transcribed from exons 1, 2, and 3 are indicated in Fig. 1A.

Knowledge of the genomic structure of *DDX1* allowed us to use the S1 protection assay, a technique that is independent of reverse transcriptase, to further define the 5'-end of the *DDX1* transcript. Poly(A)<sup>+</sup> RNAs from six *DDX1*-amplified lines (RB lines: Y79 and RB522A; NB lines: BE(2)-C, IMR-32, LA-N-1, and LA-N-5) and six nonamplified lines (RB lines: RB(E)-2 and RB412; NB lines: GOTO, NB-1, NUB-7, and SK-N-MC) were hybridized to a DNA probe that extended from position -745 in the 5'-flanking *DDX1* DNA to position +164 in exon 1. This DNA probe was labeled at position +164 as indicated in Fig. 4. Nonhybridized DNA was digested with S1 nuclease, and the sizes of the protected fragments were analyzed on a denaturing polyacrylamide gel. Bands of 150–153 nt were observed in lane 2 (RB522A), lane 5 (BE(2)-C), and lane 8 (LA-N-1) with bands of much weaker intensity in lane 7 (IMR-32) (Fig. 4). Specific bands were not detected in either *DDX1*-amplified Y79 and LA-N-5 or the nonamplified lines. Although the sizes of the S1 protected bands in RB522A, BE(2)-C, and LA-N-1 were 5 and 11 nt shorter than predicted based on RACE and primer extensions, respectively, there was general agreement with all three techniques regarding the location of the *DDX1* transcription initiation site (Fig. 1A). The smaller S1 nuclease protected products could have arisen as the result of S1 digestion of the 5'-end of the RNA:DNA heteroduplex because of its relatively high rU:dA content (45).

Identification of the same transcription initiation site in three *DDX1*-amplified lines suggests that this represents the bona fide start site of *DDX1* transcription. However, it was not clear why this start site was either very weak or not detected in three other amplified lines. To determine whether the 5'-end of exon 1 is transcribed in all *DDX1*-amplified lines, we carried out a direct analysis of the 5'-end of the *DDX1* transcript by Northern blotting. Two probes were used for this analysis: the 5' probe contained a 160-bp fragment from bp 1 to 160 (5'-half of exon 1), and the 3' probe contained a 260-bp fragment from bp 160 to 420 (3'-half of exon 1 as well as exons 2 and 3) (Fig. 1A). With the 3' probe, we obtained bands of similar size and intensity in four *DDX1*-amplified lines (RB522A, BE(2)-C, IMR-32, and LA-N-5). Band intensity was somewhat weaker in Y79 and stronger in LA-N-1 in comparison with the other lines (Fig. 5). No signal was detected in the non-*DDX1*-amplified line RB412. With the 5' probe, a relatively strong signal was observed in RB522A, BE(2)-C, and LA-N-1, while a considerably weaker

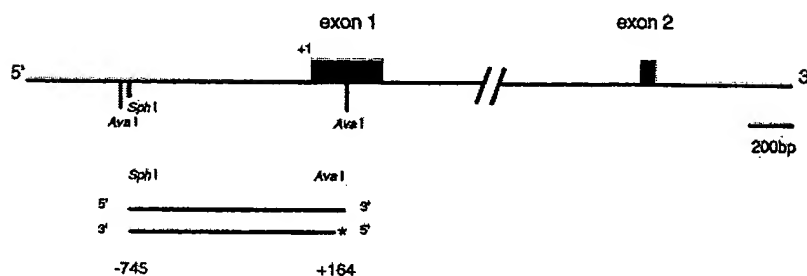
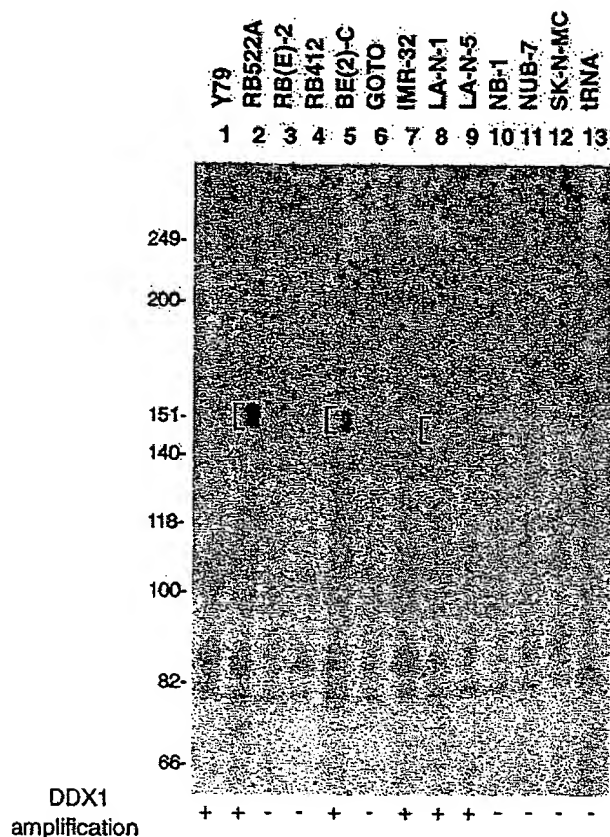
but readily apparent signal was detected in Y79, IMR-32, and LA-N-5. The signal obtained with actin indicates that, with the exception of LA-N-1, similar amounts of RNA were loaded in each lane and that the RNA was not degraded. These results indicate that at least a portion of the 160-bp 5'-end of exon 1 is transcribed in all *DDX1*-amplified lines.

Based on primer extension, S1 nuclease protection assay, Northern blot analysis and the sequencing of the RACE products, we conclude that the *DDX1* transcript is 2.7 kb with an open reading frame spanning nucleotides 295–2515 encoding a predicted protein of 740 amino acids with an estimated molecular weight of 82.4 (Fig. 1B). An in frame stop codon is located 123 nt upstream of the predicted translation initiation site, at positions 172–174. The first in frame methionine following the stop codon is in agreement with the Kozak consensus sequence (46). Furthermore, the predicted start methionine codon for human *DDX1* corresponds perfectly with that of *Drosophila* *DDX1* (47). A stop codon is located 15 nt upstream of the initiation codon in *Drosophila* *DDX1*.

**Analysis of *DDX1* Protein Levels in Neuroblastoma and Retinoblastoma**—We and others have previously shown that there is a good correlation between gene copy number and RNA levels in *DDX1*-amplified RB and NB cell lines (37, 38). To determine whether the correlation extends to *DDX1* protein levels, we prepared antiserum to two nonoverlapping recombinant *DDX1* proteins. First, we prepared a C terminus recombinant protein construct by inserting a 1.8-kb *EcoRI* fragment from bp 848 to 2668 (amino acids 185–740) (Fig. 1B) into the pMAL-c2 expression vector. Recombinant protein expression was induced with isopropyl-1-thio- $\beta$ -D-thiogalactoside, and the 110-kDa maltose-binding protein-*DDX1* fusion product was purified by affinity chromatography using amylose resin, followed by electrophoresis on a SDS-PAGE gel after cleaving the maltose-binding protein fusion partner with factor Xa. Second, we prepared an N terminus construct by ligating a DNA fragment from bp 268 to 851 (amino acids 1–186) into pGEX-4T2. The 50-kDa glutathione S-transferase-*DDX1* fusion protein was purified by affinity chromatography on a glutathione column. This N terminus fusion protein contains only the first of the eight conserved motifs found in all DEAD box proteins, while the C terminus fusion protein includes the remaining seven motifs.

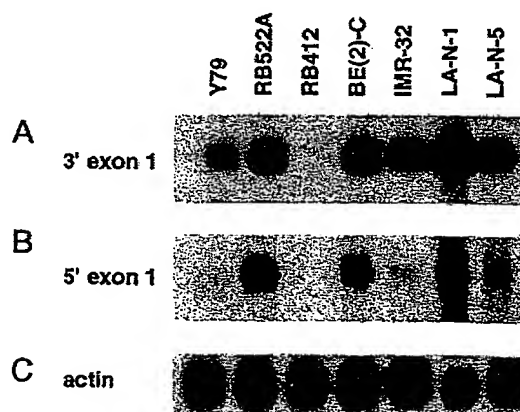
We measured *DDX1* protein levels in total cell extracts of three RB and 10 NB cell lines. Using antiserum to the N terminus fusion protein, we observed a strong signal in all *DDX1*-amplified cell lines: the RB cell lines Y79 (lane 1) and RB522A (lane 2) and the NB cell lines BE(2)-C (lane 4), IMR-32 (lane 6), LA-N-1 (lane 8), and LA-N-5 (lane 9) (Fig. 6). Two bands were observed in the majority of extracts. Of the amplified lines, Y79 produced the weakest signal, with the most intense signal observed in LA-N-1. There was an excellent correlation with *DDX1* protein and mRNA levels in these cell lines, with lower levels of *DDX1* mRNA observed in Y79 and higher levels in LA-N-1 (Fig. 7A). As shown in Fig. 7B, this correlation extended to *DDX1* gene copy number. No gross DNA rearrangements were seen in the *DDX1*-amplified lines; however, three small bands of altered size were observed in the RB412 lane. Although the nature of the DNA alteration is not known, it is noteworthy that *DDX1* transcript levels in RB412

**FIG. 4. S1 nuclease mapping of the 5'-end of the *DDX1* transcript.** Two  $\mu$ g of poly(A)<sup>+</sup> RNA from four RB lines (*DDX1*-amplified Y79 and RB522A and nonamplified RB(E)-2 and RB412), eight NB lines (*DDX1*-amplified BE(2)-C, IMR-32, LA-N-1, and LA-N-5 and nonamplified GOTO, NB-1, NUB-7, and SK-N-MC), and tRNA as a negative control were hybridized to a *Sph*I–*Ava*I fragment labeled at the *Ava*I site with [ $\gamma$ -<sup>32</sup>P]ATP and polynucleotide kinase. Bands of 150–153 nt are shown in lanes 2 (RB522A), 5 (BE(2)-C), and 8 (LA-N-1) with much weaker bands in lane 7 (IMR-32). A map of the probe indicating the transcription initiation site identified by primer extension (+1), the labeling site (\*), and exons 1 and 2, is shown at the bottom.



are extremely low (Fig. 7A) and that the top DDX1 protein band in RB412 cell extracts is smaller in size than the top band from the other cell extracts (Fig. 6).

Two DDX1 protein bands were present in most of the lanes in Fig. 6. The same two bands were detected with antiserum to the C terminus of the DDX1 protein, as well as a third band at ~60 kDa (data not shown). There was no variation in the intensity of the 60-kDa band in *DDX1*-amplified and nonamplified cell extracts. The 60-kDa band probably represents another member of the DEAD box protein family, because the C terminus DDX1 protein used to prepare this antiserum contained seven of the eight conserved motifs found in all DEAD box proteins. To obtain an estimate of the size of the two DDX1 bands, we ran cellular extracts from RB522A on a 7% SDS-PAGE gel with the BenchMark protein ladder (Life Technologies, Inc.). The size of the DDX1 protein was determined using the Alpha Imager 2000 documentation and analysis system for molecular weight calculation. Based on this analysis, the estimated molecular mass of the top band is 89.5 kDa, while that of the bottom band is 83.5 kDa. The 84-kDa band may represent the unmodified product encoded by the *DDX1* transcript (capable of encoding a protein with a predicted molecular mass of 82.4 kDa), while the top band may represent post-translational modification of DDX1 protein (e.g. phosphorylation). Another possibility is that the top band represents intact DDX1



**FIG. 5. Northern blot analysis of the 5'-end of the *DDX1* transcript.** Two  $\mu$ g of poly(A)<sup>+</sup> RNA isolated from *DDX1*-amplified Y79, RB522A, BE(2)-C, IMR-32, LA-N-1, and LA-N-5 and nonamplified RB412 were electrophoresed in a 1.5% agarose-formaldehyde gel. The RNA was transferred to a nitrocellulose filter and sequentially hybridized with a 260-bp fragment from *DDX1* cDNA from bp +160 to +420 (3'-end of exon 1 as well as exons 2 and 3) (A), a 160-bp fragment from *DDX1* cDNA from bp +1 to +160 (5'-end of exon 1) (B), and actin cDNA (C). The DNA was labeled with [<sup>32</sup>P]dCTP by nick translation. The blots were hybridized and washed under high stringency.

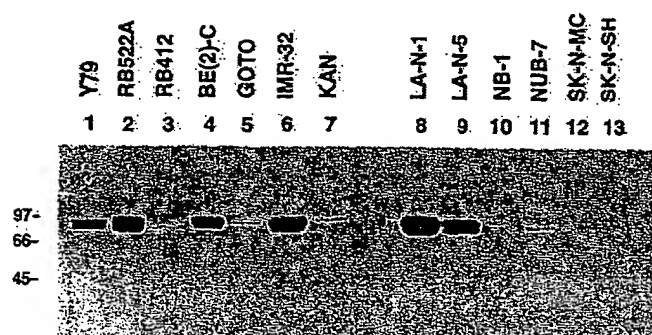


FIG. 6. DDX1 protein expression in RB and NB cell lines. Western blots were prepared using total cellular extracts from three RB (Y79, RB522A, and RB412) and 10 NB cell lines (BE(2)-C, GOTO, IMR-32, KAN, LA-N-1, LA-N-5, NB-1, NUB-7, SK-N-MC, and SK-N-SH). The lines that are amplified for the *DDX1* gene are Y79, RB522A, BE(2)-C, IMR-32, LA-N-1, and LA-N-5. Twenty  $\mu$ g of protein were loaded in each lane and electrophoresed in a 10% SDS-PAGE gel. DDX1 was detected using a 1:5000 dilution of the antiserum to the amino terminus of DDX1 protein. Size markers in kilodaltons are indicated on the side.

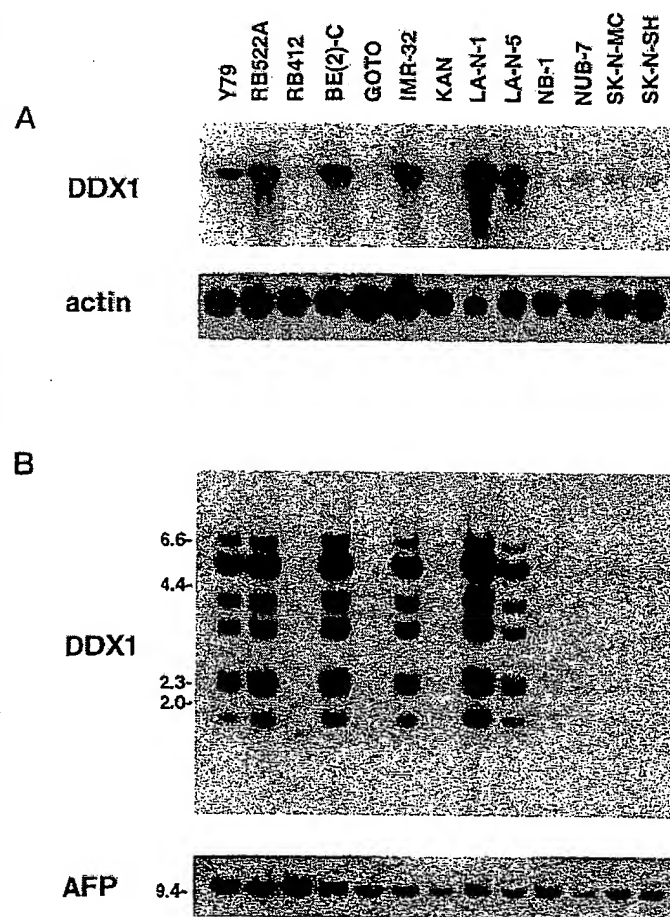


FIG. 7. Northern and Southern blot analyses of *DDX1* in RB and NB cell lines. A, 2  $\mu$ g of poly(A)<sup>+</sup> RNA were loaded in each lane, electrophoresed in a 1.5% agarose-formaldehyde gel, and transferred to a nitrocellulose filter. The filter was first hybridized to a <sup>32</sup>P-labeled 1.6-kb *DDX1* cDNA (clone 1042) (21), stripped, and rehybridized to actin DNA. B, 10  $\mu$ g of genomic DNA from each of the indicated cell lines were digested with *Eco*RI, electrophoresed in a 1% agarose gel, and transferred to a nitrocellulose filter. The filter was hybridized to <sup>32</sup>P-labeled clone 1042 *DDX1* cDNA, stripped, and rehybridized with labeled  $\alpha$ -fetoprotein cDNA. Markers (in kilobase pairs) are indicated on the side.

and the lower band is a specific truncated or degradation product of DDX1. Yet a third possibility is that the two bands represent the products of differentially spliced transcripts or

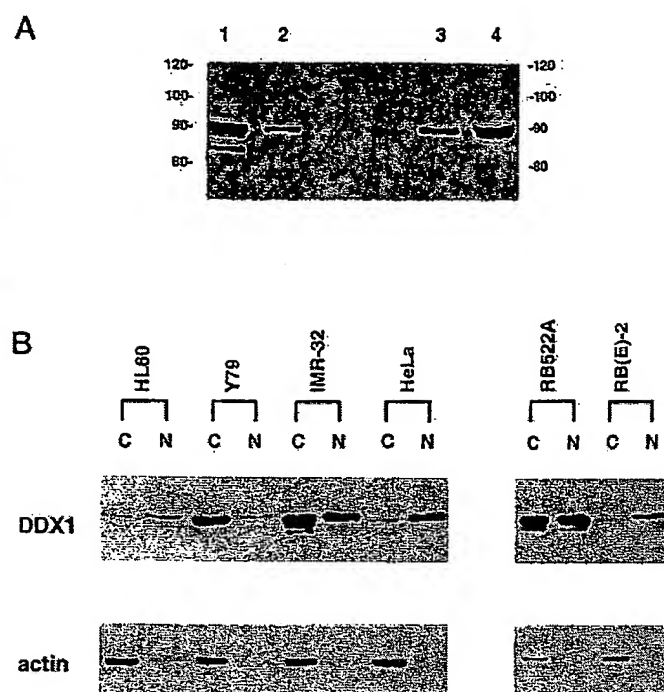


FIG. 8. Distribution of DDX1 in the nucleus and cytoplasm. A, cytosolic and nuclear extracts were prepared from RB522A and electrophoresed in a 7% SDS-PAGE gel. Cytosolic extracts were loaded in lanes 1 (20  $\mu$ g of protein) and 2 (10  $\mu$ g), while nuclear extracts were loaded in lanes 3 (10  $\mu$ g) and 4 (20  $\mu$ g). DDX1 was visualized using a 1:5000 dilution of the antiserum to the N terminus. The BenchMark protein ladder size markers (kilodaltons) are indicated on the left. B, cytosolic and nuclear extracts were prepared from HL60, Y79, IMR-32, HeLa, RB522A, and RB(E)-2 and electrophoresed in an 8% SDS-PAGE gel. Twenty  $\mu$ g of proteins were loaded in each lane marked C (cytosolic) and N (nuclear). DDX1 was visualized using a 1:5000 dilution of the antiserum to the N terminus. Actin levels were analyzed using a 1:200 dilution of anti-actin antibody (Santa Cruz Biotechnology).

different translation initiation sites. However, the lack of any obvious differences in *DDX1* transcript sizes in the three RB and 10 NB lines analyzed in Fig. 7A does not support the latter possibility (e.g. compare the *DDX1* transcript size in NUB-7 (which produces the lower DDX1 protein band) and in NB-1 (which produces the higher DDX1 protein band)).

**Subcellular Localization of DDX1 Protein**—DEAD box proteins have been implicated in a variety of cellular functions including RNA splicing in the nucleus, translation initiation in the cytoplasm, and ribosome assembly in the nucleolus. To obtain an indication of the possible role of DDX1, we studied its subcellular location. Nuclear and cytosolic extracts were prepared from *DDX1*-amplified RB522A and run on a 7% SDS-PAGE gel. Although there was more DDX1 protein in the cytosol than in the nucleus on a per cell basis, the proportion of DDX1 protein relative to total protein was similar in both cellular compartments (Fig. 8A). Both the 90- and 84-kDa bands were present in cytosol and nuclear extracts, although the bottom band was more readily apparent in the cytosol. By running the gel for an extended period of time (twice as long as usual), we were able to detect an additional weak band at ~88 kDa in both nuclear and cytosolic extracts.

To determine whether DDX1 consistently localizes to both the cytoplasm and nucleus, we prepared cytosol and nuclear extracts from two additional *DDX1*-amplified lines, Y79 and IMR-32, as well as from nonamplified RB(E)-2, HL60, and HeLa. DDX1 protein was found in both the nucleus and cytoplasm of IMR-32, primarily in the cytoplasm of Y79, and mainly in the nucleus of the three nonamplified lines (Fig. 8B). In addition, DDX1 was almost exclusively found in nuclear

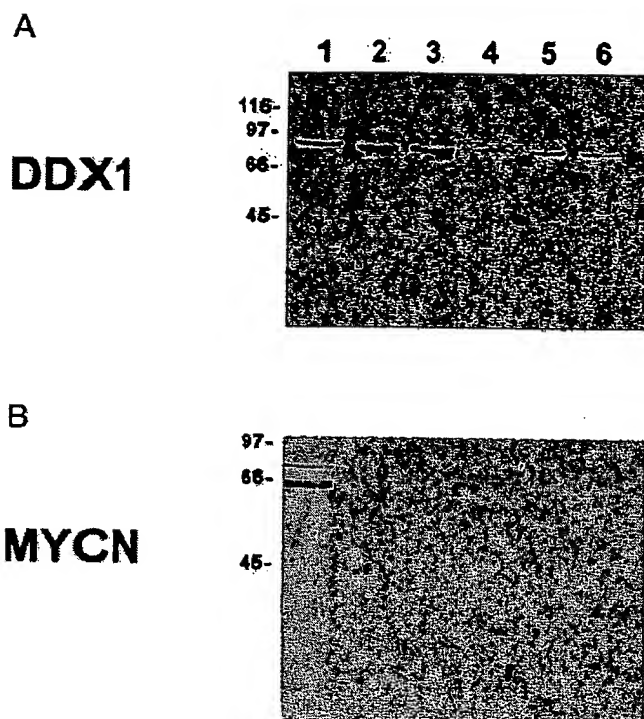


FIG. 9. Subcellular location of DDX1 protein. RB522A cells were fractionated into nuclear (lane 1), S100 and S4 cytosol (lanes 2 and 3), P2 membrane (lane 4), P3 membrane (lane 5), and P4 membrane (lane 6) fractions. Twenty  $\mu$ g of protein were loaded in each lane and run on a 10% SDS-PAGE gel. A, DDX1 protein was detected using a 1:5000 dilution of the antiserum to the N terminus of DDX1. B, MYCN protein was detected using a commercially available antibody at a 1:200 dilution. Size markers (kilodaltons) are indicated on the side.

extracts prepared from normal GM38 fibroblasts (data not shown). We used anti-actin antibody to ensure that our nuclear and cytosolic extracts were not cross-contaminated (Fig. 8B).

We next carried out a more detailed analysis of DDX1 subcellular location using two different approaches: (i) fractionation of cellular components into nuclei; S100 or S4 cytosol (containing soluble cytoplasmic components, including 40 S ribosomes); P2 (heavy mitochondria, plasma membrane fragments plus material trapped by these membranes); P3 (mitochondria, lysosomes, peroxisomes, Golgi membranes, some rough endoplasmic reticulum); and P4 (microsomes from smooth and rough endoplasmic reticulum, Golgi and plasma membranes) (43); and (ii) immunogold electron microscopy. The DDX1-amplified RB522A cell line was used for both experiments. The fractionation procedures indicate that DDX1 is mainly in the nucleus and in the cytosol (S4 and S100 fractions) of RB522A cells (Fig. 9A). As a control, we used anti-human MYCN antibody to determine the location of MYCN (also amplified in RB522A) in our subcellular fractions. As shown in Fig. 9B, MYCN was primarily found in the nucleus, as one would expect of a transcription factor.

For the electron microscopy analysis, antiserum to the N terminus of DDX1 was coupled to protein A gold particles, and the distribution of DDX1 was examined in RB522A cells fixed in paraformaldehyde and glutaraldehyde. DDX1 was present in both the cytoplasm and nucleus (data not shown). There was no association with either cell organelles or with nuclear or plasma membranes.

#### DISCUSSION

There are presently few clues as to the function of DDX1 in normal and cancer cells. Our earlier data indicate that DDX1 mRNA is present at higher levels in fetal tissues of neural origin (retina and brain) compared with other fetal tissues (21).

There may therefore be a requirement for elevated levels of this putative RNA helicase for the efficient production or processing of neural specific transcripts. A role in cancer formation or progression is an intriguing possibility, because overexpression of an RNA unwinding protein could affect the secondary structure of RNAs in such a way as to alter the expression of specific proteins in tumor cells. DDX1 is co-amplified with MYCN in a subset of RB and NB cell lines and tumors (37–39). MYCN amplification is common in stage IV NB tumors and is a well documented indicator of poor prognosis. A general trend toward a poorer clinical prognosis is observed when both the MYCN and DDX1 genes are amplified compared with when only MYCN is amplified (38, 39), suggesting a possible role for DDX1 in NB tumor formation or progression.

It is generally accepted that co-amplified genes are not overexpressed unless they provide a selective growth advantage to the cell (48, 49). For example, although ERBA is closely linked to ERBB2 in breast cancer and both genes are commonly amplified in these tumors, ERBA is not overexpressed (48). Similarly, three genes mapping to 12q13–14 (CDK4, SAS, and MDM2) are overexpressed in a high percentage of malignant gliomas showing amplification of this chromosomal region, while other genes mapping to this region (GADD153, GLI, and A2MR) are rarely overexpressed in gene-amplified malignant gliomas (50, 51). The first three genes are probably the main targets of the amplification process, while the latter three genes are probably incidentally included in the amplicons. The data shown here indicate that DDX1 is overexpressed at both the protein and RNA levels in DDX1-amplified RB and NB cell lines and that there is a strong correlation between DDX1 gene copy number, DDX1 RNA levels, and DDX1 protein levels in these lines. Our results are therefore consistent with DDX1 overexpression playing a positive role in some aspect of NB and RB tumor formation or progression. Recently, Weiss *et al.* (52) have shown that transgenic mice that overexpress MYCN develop NB tumors several months after birth. They conclude that MYCN overexpression can contribute to the initiation of tumorigenesis but that additional events are required for tumor formation. Amplification of DDX1 may represent one of many alternative pathways by which a normal precursor "neuroblast" or "retinoblast" cell gains malignant properties.

The function of the majority of tissue-specific or developmentally regulated DEAD box genes remains unknown. However, some members of this protein family have been either directly or indirectly implicated in tumorigenesis. For example, the p68 gene has been found to be mutated in the ultraviolet light-induced murine tumor 8101 (53), while DDX6 (also known as RCK or p54) is encoded by a gene located at the breakpoint of the translocation involving chromosomes 11 and 14 in a cell line derived from a B-cell lymphoma (54, 55). Similarly, the production of a chimeric protein between DDX10 and the nucleoporin gene NUP98 has been proposed to be involved in the pathogenesis of a subset of myeloid malignancies with inv(11) (p15q22) (56). Interestingly, Grandori *et al.* (57) have shown that MYCC interacts with a DEAD box gene called MrDb, suggesting that the transcription of some DEAD box genes could be regulated through interaction with members of the MYC family. Future work will involve determining whether DDX1 represents another member of the DEAD box family with a role in the tumorigenic process.

DEAD box proteins have been implicated in translation initiation, RNA splicing, RNA degradation, and RNA stability (3, 18, 19). We carried out subcellular localization studies in an attempt to obtain a general indication of the function of DDX1. We found DDX1 protein in both the cytoplasm and nucleus of DDX1-amplified NB and RB lines. In contrast, DDX1 was

mainly located in the nucleus of nonamplified cell lines and normal fibroblast cultures. DDX1 was not associated with cellular organelles or with membranes based on immunoelectron microscopy. We therefore propose that the primary role of DDX1 is in the nucleus. The presence of DDX1 in the cytoplasm of DDX1-amplified cells may indicate that the amount of DDX1 protein that is allowed in the nucleus is tightly regulated. Alternatively, DDX1 may play a dual role in the nucleus and cytoplasm of DDX1-amplified cells.

An important component of our analysis was to identify the translation and transcription initiation sites of DDX1. We used a combination of techniques to identify the transcription start site: screening of RB and fetal brain libraries, RACE, primer extension, genomic DNA sequencing, S1 nuclease mapping, and Northern blot analysis using probes to the predicted 5'-end of the transcript. The transcription start site identified using these techniques is located ~300 nt upstream of the predicted translation initiation codon and was readily detected in three DDX1-amplified lines and barely detectable in a fourth amplified line. The 5'-untranslated region as well as the first in frame methionine are encoded within the first exon of DDX1. An in frame stop codon is located 123 nt upstream of the predicted initiation codon. We were unable to identify the transcription initiation site of DDX1 in two of the six amplified lines tested as well as in nonamplified lines. Although it remains possible that there are different transcription start sites in different cell lines, detection of lower levels (rather than the absence) of the 5'-most 160 nt of the DDX1 transcript in IMR-32, Y79, and LA-N-5 compared with RB522A, BE(2)-C, and LA-N-1 supports a quantitative rather than a qualitative difference in the 5'-end of this transcript in these cells. Our results suggest that the 5'-end of DDX1 mRNA is rarely intact, even in mRNA preparations that otherwise appear to be of high quality based on analysis of control transcripts. The 5'-end of DDX1 mRNA may therefore be especially susceptible to degradation, perhaps because of its sequence and/or secondary structure.

In conclusion, we have mapped the 5'-end of the 2.7-kb DDX1 transcript and have identified the predicted translation initiation site of DDX1 protein. We have found that DDX1-amplified RB and NB tumor lines overexpress DDX1 protein and that there is a good correlation between gene copy number and both transcript and protein levels in these cells. We have shown that DDX1 protein is primarily located in the nucleus of cells that are not DDX1-amplified. In contrast, DDX1 is present in both the nucleus and cytoplasm of DDX1-amplified NB and RB lines. A cytoplasmic location in DDX1-amplified lines may indicate that the amount of nuclear DDX1 is tightly regulated or that DDX1 plays a dual role in the cytoplasm and nucleus of these cells.

**Acknowledgments**—We thank Walter Dixon, Brenda Gallie, Ajay Pandita, Jeremy Squire, and Herman Yeger for the neuroblastoma and retinoblastoma cell lines. We thank Halyna Marusyk for carrying out the electron microscopy analyses. We are grateful to Randy Anderson for expert help in the preparation of the manuscript and to Stacey Hume for helpful discussions.

#### REFERENCES

- Linder, P., Lasko, P. F., Ashburner, M., Leroy, P., Nielsen, P. J., Nishi, K., Schnier, J., and Slonimski, P. P. (1989) *Nature* 337, 121–122
- Wasserman, D. A., and Steitz, J. A. (1991) *Nature* 349, 463–464
- Schmid, S. R., and Linder, P. (1992) *Mol. Microbiol.* 6, 283–292
- Rozen, F., Edery, I., Meerovitch, K., Dever, T. E., Merrick, W. C., and Sonenberg, N. (1990) *Mol. Cell. Biol.* 10, 1134–1144
- Hirling, H., Scheffner, M., Restle, T., and Stahl, H. (1989) *Nature* 339, 562–564
- Liang, L., Diehl-Jones, W., and Lasko, P. (1994) *Development* 120, 1201–1211
- Gururajan R., Mathews, L., Longo, F., and Weeks, D. L. (1994) *Proc. Natl. Acad. Sci. U. S. A.* 91, 2056–2060
- Dalbadie-McFarland, G., and Abelson, J. (1990) *Proc. Natl. Acad. Sci. U. S. A.* 87, 4236–4240
- Jamieson, D. J., Rahe, B., Pringle, J., and Beggs, J. D. (1991) *Nature* 349, 715–717
- Strauss, E. J., and Guthrie, C. (1994) *Nucleic Acids Res.* 22, 3187–3193
- Nishi, K., Morel-Deville, F., Hershey, J. W. B., Leighton, T., and Schnier, J. (1988) *Nature* 336, 496–498
- Leroy, P., Alzari, P., Sassoon, D., Wolgemuth, D., and Fellous, M. (1989) *Cell* 57, 549–559
- Hay, B., Jan, L. Y., and Jan, Y. N. (1988) *Cell* 55, 577–587
- De Valoir, T., Tucker, M., Belikoff, E. J., Camp, L. A. Bolduc, C., and Beckingham, K. (1991) *Proc. Natl. Acad. Sci. U. S. A.* 88, 2113–2117
- Ford, M. J., Anton, I. A., and Lane, D. P. (1988) *Nature* 332, 736–738
- Iggo, R. D., and Lane, D. P. (1989) *EMBO J.* 8, 1827–1831
- Buelt, M. K., Glidden, B. J., and Storm, D. R. (1994) *J. Biol. Chem.* 269, 29367–29370
- Iost, I., and Dreyfus, M. (1994) *Nature* 372, 193–196
- Py, B., Higgins, C. F., Krisch, H. M., and Carpousis, A. J. (1996) *Nature* 381, 169–172
- Missel, A., Souza, A. E., Nörskau, G., and Göringer, H. U. (1997) *Mol. Cell. Biol.* 17, 4895–4903
- Godbout, R., and Squire, J. (1993) *Proc. Natl. Acad. Sci. U. S. A.* 90, 7578–7582
- Kiledjian, M., and Dreyfuss, G. (1992) *EMBO J.* 11, 2655–2664
- Godbout, R., Hale, M., and Bisgrove, D. (1994) *Gene (Amst.)* 138, 243–245
- Blackwood, E. M., and Eisenman, R. N. (1991) *Science* 251, 1211–1217
- Amati, B., Dalton, S., Brooks, M. W., Littlewood, T. D., Evan, G. I., and Land, H. (1992) *Nature* 359, 423–426
- Brodeur, G. M., Seeger, R. C., Schwab, M., Varmus, H. E., and Bishop, J. M. (1984) *Science* 224, 1121–1124
- Seeger, R. C., Brodeur, G. M., Sather, H., Dalton, A., Siegel, S. E., Wong, K. Y., and Hammond, D. (1985) *N. Engl. J. Med.* 313, 1111–1116
- Cohn, S. L., Salwen, H., Quasney, M. W., Ikegaki, N., Cowan, J. M., Herst, C. V., Sharon, B., Kennett, R. H., and Rosen, S. T. (1991) *Frog. Clin. Biol. Res.* 366, 21–27
- Cowell, J. K. (1982) *Annu. Rev. Genet.* 16, 21–59
- Zehnauer, B. A., Small, D., Brodeur, G. M., Seeger, R., and Vogelstein, B. (1988) *Mol. Cell. Biol.* 8, 522–530
- Akiyama, K., and Nishi, Y. (1991) *Nucleic Acids Res.* 19, 6887–6894
- Schneider, S. S., Hiemstra, J. L., Zehnauer, B. A., Taillon-Miller, P., Le Paslier, D., Vogelstein, B., and Brodeur, G. M. (1992) *Mol. Cell. Biol.* 12, 5563–5570
- Amler, L. C., Schürmann, J., and Schwab, M. (1996) *Genes Chromosomes Cancer* 15, 134–137
- Kuroda, H., White, P. S., Sulman, E. P., Manohar, C., Reiter, J. L., Cohn, S. L., and Brodeur, G. M. (1996) *Oncogene* 13, 1561–1565
- Noguchi, Y., Akiyama, K., Yokoyama, M., Kanda, N., Matsunaga, T., and Nishi, Y. (1996) *Genes Chromosomes Cancer* 15, 129–133
- Pandita, A., Godbout, R., Zielenska, M., Thorner, P., Bayani, J., and Squire, J. A. (1997) *Genes Chromosomes Cancer* 20, 243–252
- Manohar, C. F., Salwen, H. R., Brodeur, G. M., and Cohn, S. L. (1995) *Genes Chromosomes Cancer* 14, 196–203
- Squire, J. A., Thorner, P. S., Weitzman, S., Maggi, J. D., Dirks, P., Doyle, J., Hale, M., and Godbout, R. (1995) *Oncogene* 10, 1417–1422
- George, R. E., Kenyon, R. M., McGuckin, A. G., Malcolm, A. J., Pearson, A. D. J., and Lunec, J. (1996) *Oncogene* 12, 1583–1587
- Reiter, J. L., and Brodeur, G. M. (1996) *Genomics* 32, 97–103
- Favaloro, J., Treisman, R., and Kamen, R. (1980) *Methods Enzymol.* 65, 718–745
- Dignam, J. D. (1990) *Methods Enzymol.* 182, 194–203
- Graham, J. (1984) in *Centrifugation: A Practical Approach* (Rickwood, D. ed.) 2nd Ed., pp. 161–182, IRL Press, Oxford
- Godbout, R., Marusyk, H., Bisgrove, D., Dabbagh, L., and Poppema, S. (1995) *Exp. Eye Res.* 60, 645–657
- Miller, K. G., and Sollner-Webb, B. (1981) *Cell* 27, 165–174
- Kozak, M. (1988) *Nucleic Acids Res.* 15, 8125–8132
- Rafti, F., Scarvelis, D., and Lasko, P. F. (1996) *Gene (Amst.)* 171, 225–229
- Van de Vijver, M., van de Berselaar, R., Devilee, P., Cornelisse, C., Peterse, J., and Nusse, R. (1987) *Mol. Cell. Biol.* 7, 2019–2023
- Gaudray, P., Szegetowski, P., Escot, C., Birnbaum, D., and Theillet, C. (1992) *Mutat. Res.* 276, 317–328
- Forus, A., Florence, V. A., Maelandsmo, G. M., Meltzer, P. S., Fodstad, O., and Myklebost, O. (1993) *Cell Growth Differ.* 4, 1065–1070
- Reifenberger, G., Reifenberger, J., Ichimura, K., Meltzer, P. S., and Collins, V. P. (1994) *Cancer Res.* 54, 4299–4303
- Weiss, W. A., Aldape, K., Mohapatra, G., Feuerstein, B. G., and Bishop, J. M. (1997) *EMBO J.* 16, 2985–2995
- Dubey, B. P., Hendrickson, R. C., Meredith, S. C., Siegel, C. T., Shabanowitz, J., Skipper, J. C. A., Engelhard, V. H., Hunt, D. F., and Schreiber, H. (1997) *J. Exp. Med.* 185, 695–705
- Akao, Y., Seto, M., Yamamoto, K., Iida, S., Nakazawa, S., Inazawa, J., Abe, T., Takahashi, T., and Ueda, R. (1992) *Cancer Res.* 52, 6083–6087
- Lu, D., and Yunis, J. J. (1992) *Nucleic Acids Res.* 20, 1967–1972
- Arai, Y., Hosoda, F., Kobayashi, H., Arai, K., Hayashi, Y., Kamada, N., Kaneko, Y., and Ohki, M. (1997) *Blood* 89, 3936–3944
- Grandori, C., Mac, J., Siébelt, F., Ayer, D. E., and Eisenman, R. N. (1996) *EMBO J.* 15, 4344–4357



ORIGINAL ARTICLE

## Identification of putative oncogenes in lung adenocarcinoma by a comprehensive functional genomic approach

R Li<sup>1</sup>, H Wang<sup>1</sup>, BN Bekele<sup>2</sup>, Z Yin<sup>3</sup>, NP Caraway<sup>1</sup>, RL Katz<sup>1</sup>, SA Stass<sup>4</sup> and F Jiang<sup>4</sup>

<sup>1</sup>Department of Pathology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA; <sup>2</sup>Department of Biostatistics & Applied Mathematics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA; <sup>3</sup>Department of Internal Medicine, The University of Texas Houston Medical School, Houston, TX, USA and <sup>4</sup>Department of Pathology, The University of Maryland School of Medicine, Baltimore, MD, USA

Amplification and overexpression of putative oncogenes confer growth advantages for tumor development. We used a functional genomic approach that integrated simultaneous genomic and transcript microarray, proteomics, and tissue microarray analyses to directly identify putative oncogenes in lung adenocarcinoma. We first identified 183 genes with increases in both genomic copy number and transcript in six lung adenocarcinoma cell lines. Next, we used two-dimensional polyacrylamide gel electrophoresis and mass spectrometry to identify 42 proteins that were overexpressed in the cancer cells relative to normal cells. Comparing the 183 genes with the 42 proteins, we identified four genes – *PRDX1*, *EEF1A2*, *CALR*, and *KCIP-1* – in which elevated protein expression correlated with both increased DNA copy number and increased transcript levels (all  $r > 0.84$ , two-sided  $P < 0.05$ ). These findings were validated by Southern, Northern, and Western blotting. Specific inhibition of *EEF1A2* and *KCIP-1* expression with siRNA in the four cell lines tested suppressed proliferation and induced apoptosis. Parallel fluorescence *in situ* hybridization and immunohistochemical analyses of *EEF1A2* and *KCIP-1* in tissue microarrays from patients with lung adenocarcinoma showed that gene amplification was associated with high protein expression for both genes and that protein overexpression was related to tumor grade, disease stage, Ki-67 expression, and a shorter survival of patients. The amplification of *EEF1A2* and *KCIP-1* and the presence of overexpressed protein in tumor samples strongly suggest that these genes could be oncogenes and hence potential targets for diagnosis and therapy in lung adenocarcinoma. *Oncogene* (2006) 25, 2628–2635. doi:10.1038/sj.onc.1209289; published online 12 December 2005

**Keywords:** lung cancer; microarrays; proteomics; tissue microarray

### Introduction

In lung adenocarcinoma, as in other types of cancer, gene amplification and the consequent overexpression of the amplified oncogene play an important role in the development of tumors, because their overexpression confers a growth advantage. The ability to identify putative oncogenes that are activated during tumorigenesis could facilitate the choice of molecular genetic targets for diagnosis and therapy of the disease. This concept has been exemplified by *HER-2*, which was first found to be amplified in neuroblastomas and subsequently shown to be associated with poor prognosis in breast cancer (Ross and Fletcher, 1999). Now, *HER-2* aberrations are used as a predictor of response to therapy, and treatment of *HER-2*-positive breast cancer with the monoclonal anti-*HER-2* antibody trastuzumab has been shown to improve prognosis (Ross and Fletcher, 1999). Emerging evidence of common amplicons in lung adenocarcinomas (Luk *et al.*, 2001; Jiang *et al.*, 2004; Tonon *et al.*, 2005) suggests that additional oncogenes remain to be identified; however, conventional techniques are ineffective in pinpointing such oncogenes. Parallel measurement of DNA copy number and mRNA levels in cDNA microarrays permits changes in copy number to be compared with transcription levels on a gene-by-gene basis to generate lists of candidate genes within the defining amplicons (Hyman *et al.*, 2002; Pollack *et al.*, 2002). However, use of transcript patterns does not allow assessment of the expression of protein products or identification of proto-oncogenes. Another approach, identifying differentially expressed proteins by proteomic analysis and then comparing the proteins present with mRNA expression in cDNA microarrays from the same specimens, can clarify the extent to which changes in transcript patterns reflect changes in their cognate proteins and post-transcriptional mechanisms (Chen *et al.*, 2002), but this approach cannot be used to identify oncogenes driven by extensive increases of their gene copy number. Moreover, using individual microarrays or proteomic approaches alone cannot distinguish the cancer-driving oncogenes that directly propel tumor progression from the larger number of passenger genes that may be concurrently over-represented but are not biologically relevant in tumor development.

Correspondence: Assistant Professor F Jiang, Department of Pathology, The University of Maryland School of Medicine, 10 South Pine Street, MSTF 7th floor, Baltimore, MD 21201-1192, USA.  
E-mail: fjiang@som.umaryland.edu  
Received 27 July 2005; revised 19 October 2005; accepted 27 October 2005; published online 12 December 2005

In this study, we used a comprehensive approach that integrated simultaneous comparative genomic hybridization (CGH) and transcript microarray with proteomic analyses of six lung adenocarcinoma cell lines. We directly and specifically identified four putative oncogenes that could have been activated through amplification and consequent elevation of transcript expression. We used small interfering RNA (siRNA) to inhibit the expression of two of these four genes in the lung cancer cell lines, which further implicated them in oncogenesis. We then explored the clinical significance of these findings by assessing the expression of these two genes in tissue microarrays of human lung cancer specimens. Our findings underscore the power of integrated functional genomic analyses for identifying putative oncogenes in tumorigenesis; such activated genes could be useful as targets for diagnosis or therapy in lung cancer.

## Results

### *Simultaneous global genomic and transcript analyses identify 183 genes with increases in genomic copy numbers and transcript expression levels*

To identify genes in which increased DNA copy number might contribute to increased transcript in lung adenocarcinomas, first we used CGH with microarrays of six lung adenocarcinoma cell lines. We identified 587 genes showing increases in DNA copy number across all six cell lines (Supplementary Table 1S), which were distributed as 90 amplicons on all chromosomes except for chromosomes 13 and Y (Supplementary Table 2S). A subsequent transcript test with the identical arrays of the same cell lines revealed 275 genes that showed increased mRNA levels (Supplementary Table 3S). Using random permutation tests across all cancer cell lines, we identified 183 genes (31%) that showed elevated transcript levels from the 587 genes that were over-represented in the genome (Table 1), suggesting that elevated transcript levels of the 183 genes may reflect their genomic over-representation in the cancer cells. These findings are consistent with previous reports linking genomic changes with altered transcript patterns in breast cancer (Hyman *et al.*, 2002; Pollack *et al.*, 2002). However, our finding that only 31% of the genes showing increased DNA copy numbers had cognate increases in transcript expression in lung adenocarcinomas is different from the overall rates of 40–60% reported for breast cancer (Hyman *et al.*, 2002; Pollack *et al.*, 2002). This discordance may reflect methodologic differences between studies or biological differences between breast cancer and lung adenocarcinoma.

### *Proteomic analyses identify four genes for which protein abundance was associated with increases in the cognate gene and transcript levels*

Analysis of transcript patterns is insufficient for understanding the expression of protein products and the effect of genomic over-representation on the expression

of their cognate proteins. To extend these findings beyond genomic over-representation to expression of the protein products of those genes, we next assessed protein expression in the same cell lines by two-dimensional polyacrylamide gel electrophoresis (PAGE) and found that 42 different proteins, representing 42 individual genes, were significantly increased in the cancer cell lines (Table 2; Supplementary Figures 1S and 2S). Some of these proteins were identified as having multiple isoforms, and all individual isoforms exhibited increases in expression ranging from 4.6 to 12.8 times their expression in normal lung tissue cells. In comparing protein level of the 42 genes with changes in their cognate genomic and mRNA expression from the global microarray analyses, we found that four (9.5%) of those 42 genes – *PRDX1*, *EEF1A2*, *CALR*, and *KCIP-1* – showed statistically significant correlations between elevated protein expression and increases in both copy number and mRNA expression (all  $r > 0.84$ ;  $P < 0.05$ ) (Table 2) in the cancer cell lines. These findings imply that the abundance of these four proteins is attributable to the amplification and consequent elevated transcription of their cognate genes.

### *Validation of copy number, transcript, and protein expression of PRDX1, EEF1A2, CALR, and KCIP-1 in lung cancer cell lines*

To confirm our findings from the high-throughput analyses, we next used Southern, Northern, and Western blotting to assess DNA, RNA, and protein levels for the four genes identified in the six cell lines. For comparison, we arbitrarily chose one gene, *NFKB1*, in which an increase in protein level did not correlate with genetic changes. Overall, we found excellent concordance between the CGH microarray and Southern blotting analyses, transcript array and Northern blotting analyses, and proteomic and Western blotting analyses for all five genes (Figure 1). For example, *KCIP-1* showed fivefold amplification in five of the six cancer cell lines, whereas *NFKB1* showed no such increase in any of the cell lines. As for transcript expression, Northern blotting of *EEF1A2* showed high expression in five of the six cancer cell lines; again, levels of *NFKB1* transcript were not increased in any cancer cell line as compared with normal bronchial epithelial cells. The results of Western blotting were also consistent with the results of the proteomic experiments; for example, five of the cancer cell lines exhibited strong protein bands for *PRDX1* as compared with normal cells. These findings provide strong support for the validity of the results derived from the high-throughput techniques in this study.

These parallel analyses also revealed close correlations in the extent of changes in gene copies, transcript, and protein of each of the four genes in the cancer cell lines. For example, in the five cancer cell lines that showed at least fourfold increases in *EEF1A2* copy number, expression of transcript and protein was also increased by at least a factor of four as well (relative to their expression in normal cells) (Supplementary Figure 3S). The protein abundance of the four genes showing

Table 1 List of 183 genes with statistically significant correlation (0.05) between genomic copy number and transcript level

Gene symbol	Chro.	Distance from p arm of each chromosome (Mb)	$\alpha$
ENO1	1	8.5	0.0085
DDOST	1	20.1	0.0111
SFN	1	26.4	0.0113
MLP	1	32.2	0.0114
AKR1A1	1	45.4	0.0128
PRDX1	1	45.4	0.0122
UQCRH	1	46.2	0.0125
RPL7	1	96.4	0.0127
COL11A1	1	102.6	0.0129
MCL1	1	147.3	0.0222
PSMB4	1	148.1	0.0131
JTB	1	150.7	0.0134
RPS27	1	150.7	0.0135
HAX1	1	151	0.0266
MUC1	1	151.9	0.0143
CCT3	1	153.1	0.0167
CRABP2	1	153.4	0.0148
TKT	1	159.3	0.0152
ATP1B1	1	165.8	0.0234
CHIT1	1	199.7	0.0154
SNRPE	1	200.2	0.0165
YWHAQ	2	9.6	0.0159
ODC1	2	10.60	0.0119
RPL31	2	101.20	0.0161
BENE	2	110.40	0.0169
STAT1	2	191.80	0.0175
HSPD1	2	198.30	0.0277
HSPE1	2	198.30	0.0185
RPL37A	2	217.30	0.0388
IGFBP2	2	217.50	0.0189
RPS7	2	3.30	0.0193
RAB1A	2	65.30	0.0204
IGKC	2	89.00	0.0285
LTF	3	46.3	0.0455
PFN2	3	151	0.0207
KPNA4	3	161.5	0.0211
SI00P	4	6.7	0.1122
UGDH	4	39.3	0.0215
UCHL1	4	41.1	0.0222
SPPI	4	89.3	0.0227
TRIM2	4	154.7	0.0231
FGF	4	156	0.0235
SDHA	5	0.251	0.0243
PDCD6	5	0.305	0.0245
CCT5	5	10.3	0.0446
PTPRF	5	14.2	0.0248
RPL37	5	40.8	0.0251
ENC1	5	74	0.0336
QP-C	5	132.2	0.0466
SPINK1	5	147.2	0.0256
CANX	5	179.2	0.0263
SOX4	6	21.7	0.0321
HDGF	6	22.6	0.0362
RPS10	6	34.6	0.0177
RPL10A	6	35.4	0.0369
VEGF	6	43.7	0.0372
OSF-2	6	45.4	0.0173
FSCN1	7	5.3	0.0378
CYCS	7	24.9	0.0381
CBX3	7	25.9	0.0289
IGFBP3	7	45.7	0.0389
CLDN4	7	72.7	0.0403
HSPB1	7	75.5	0.0433
CALR	7	92.7	0.0425
COL1A2	7	93.6	0.0457
ATP5J2	7	98.7	0.0475
AKR1B10	7	133.6	0.0481

Table 1 (continued)

Gene symbol	Chro.	Distance from p arm of each chromosome (Mb)	$\alpha$
RPS20	8	56.7	0.0482
TCEB1	8	74.6	0.0486
LAPTM4B	8	98.5	0.0497
RPL30	8	98.7	0.0054
KCIP-1	8	101.6	0.0093
PABPC1	8	101.78	0.0119
EEF1D	8	144.4	0.0121
TSTA3	8	144.5	0.0122
RPL8	8	145.6	0.0128
TRA1	9	117.1	0.0136
RPL35	9	121.1	0.0133
HSPA5	9	121.5	0.0135
LCN2	9	124.4	0.0137
DPP7	9	133.4	0.0139
PFKP	10	3.2	0.0223
AKR1C1	10	5.1	0.0146
PLAU	10	75.6	0.0356
DSP	10	76.7	0.0289
TALDO1	11	0.434	0.0143
SLC22A1L	11	2.9	0.0151
TSSC3	11	2.9	0.0611
RPL27A	11	8.7	0.0156
ST5	11	8.8	0.0162
LDHA	11	18.5	0.0168
MDK	11	46.4	0.0162
DOC-1R	11	67.5	0.0167
MMP12	11	102.8	0.0177
HYOU1	11	118.9	0.0183
SCNN1A	12	6.3	0.0185
LDHB	12	21.7	0.0193
KRT7	12	52.3	0.0196
KRT5	12	52.6	0.0197
KRT6E	12	52.6	0.0201
ERBB3	12	56.2	0.0212
NACA	12	56.8	0.0218
TM4SF3	12	71.2	0.0401
NTS	12	86.2	0.0215
ASCL1	12	103.3	0.0219
TXNRD1	12	104.6	0.0223
CKAP4	12	106.6	0.0124
COX6A1	12	120.7	0.0435
BGN	12	122.5	0.0235
RAN	12	129.88	0.0238
RPL36A	14	48.1	0.0243
PGD	14	50.7	0.0248
THBS2	15	37.5	0.0251
TRAF4	15	38.3	0.0253
SPINT1	15	38.7	0.0254
RPL17	15	45.26	0.0411
PKM2	15	70.1	0.0258
IDH2	15	88.2	0.0211
RPL23A	16	0.377	0.0264
MSLN	16	0.753	0.0366
UBE2I	16	1.3	0.0271
RPS2	16	1.95	0.0281
CLDN9	16	3.1	0.0329
ARL6IP	16	18.7	0.0412
EIF3S8	16	28.3	0.0336
TUFM	16	28.9	0.0377
ALDOA	16	30.1	0.038
NME4	16	53.6	0.0381
GPR56	16	57.4	0.0386
CDH1	16	68.5	0.0289
NQO1	16	69.5	0.0396
SLC7A5	16	87.6	0.0397
APRT	16	88.6	0.0411
GALNS	16	88.6	0.0255
RPL13	16	89.3	0.0431
MCP	17	32.4	0.0465

Table 1 (continued)

Gene symbol	Chro.	Distance from p arm of each chromosome (Mb)	$\alpha$
ERBB2	17	35.11	0.0483
JUP	17	39.8	0.0495
CRF	17	40.39	0.0505
RPL27	17	41.1	0.0046
NME1	17	46.59	0.0082
COL1A1	17	48.6	0.0108
ABCC3	17	49.1	0.0326
NME2	17	49.6	0.0111
RPL38	17	72.7	0.0117
SMT3H2	17	73.6	0.0119
SYNGR2	17	76.6	0.0122
LGALS3BP	17	77.4	0.0127
P4HB	17	80.3	0.0126
PPAP2C	19	0.221	0.0228
GPI	19	39.55	0.0145
HPN	19	40.2	0.0129
ZNF146	19	41.4	0.0131
SPINT2	19	43.4	0.0238
PSMD8	19	43.5	0.0132
YIF1P	19	43.5	0.0135
RPS16	19	44.6	0.0144
CEACAM5	19	46.9	0.0145
CEACAM6	19	46.9	0.0143
GIPR	19	50.8	0.0259
SNRPD2	19	50.9	0.0413
KDELRL1	19	53.6	0.0152
RPL28	19	60.6	0.0156
RPS5	19	63.6	0.0267
TRIM28	19	63.7	0.0158
DAP	20	35.6	0.0166
TOP1	20	40.3	0.0172
UBE2C	20	45.1	0.0174
RPS21	20	61.6	0.0268
EEF1A2	20	62.8	0.0185
TFE3	21	42.6	0.0186
TFE1	21	42.7	0.0192
CSTB	21	44.1	0.0201
MIF	22	22.6	0.0202
XBP1	22	27.5	0.0204
PRDX4	X	22.9	0.0198
SYN1	X	46.3	0.0204
TIMP1	X	46.3	0.0209
PLP2	X	47.8	0.0212
MAGED1	X	50.3	0.0331
RPS4X	X	71	0.0124
SSR4	X	152.6	0.0232

corresponding increases in both DNA copy number and mRNA provides further evidence that these could be oncogenes, the activation of which is reflected by genomic amplification and consequent increases in transcript level in lung adenocarcinoma cell lines.

#### Specific inhibition of EEF1A2 and KCIP-1 expression by siRNAs led to decreased cell proliferation and induction of apoptosis

To further prove the oncogenic function of the identified genes in lung tumorigenesis, we used siRNAs to inhibit the endogenous expression of EEF1A2 and KCIP-1 protein in four lung cancer cell lines (H1563, H229, H522, and SK-LU). Transfection of the cancer cells with specific siRNAs reduced the level of EEF1A2 and KCIP-1 protein by 70–90% 48 h after transfection

(Supplementary Figure 4S). In contrast, EEF1A2 and KCIP-1 protein levels remained unchanged in mock-treated control cells and in cells transfected with a scrambled siRNA sequence. At 48 h after siRNA transfection, the percentage of proliferation of the transfected cancer cells was reduced to 15–30% as compared with 91–100% of cell proliferation of the same cell lines treated with PBS or scrambled siRNA (Supplementary Figure 5S). Apoptosis of siRNA-transfected cells was 27–34%, whereas only 4% of the same cell lines treated with PBS or scrambled siRNA showed apoptosis. These results strongly support an oncogenic role for the identified genes in lung cancer and confirm their potential usefulness as therapeutic targets for the disease.

#### Amplification and protein expression of KCIP-1 and EEF1A2 in lung tissue

To further validate these findings and to assess the possible clinical significance of the four potential putative oncogenes identified from the cell lines, we first applied fluorescence *in situ* hybridization and immunohistochemical analysis, in parallel, to commercially available human lung tissue microarrays (Ambion, Austin, TX, USA) to evaluate the status of two of these four genes in lung cancer tissue specimens. (Commercially available antibodies to PRDX1 or CALR were not suitable for use in immunohistochemical analysis when this report was written.) Overexpression of KCIP-1 and EEF1A2 protein in the tumors was concordant with amplification of the corresponding genes ( $P=0.0003$  for KCIP-1 and  $P=0.0011$  for EEF1A2). For example, 16 (35%) of the 46 lung adenocarcinomas in the microarray showed amplification of *KCIP-1*, and strong cytoplasmic staining for KCIP-1 protein was seen in 18 tumors (39%) (Figure 2). We next examined whether overexpression of these genes was associated with increased cell proliferation by analysing Ki-67 expression in contiguous sections of the tissue microarrays. Positive Ki-67 expression was found to correlate with positive expression of both KCIP-1 ( $P=0.02$ ) and EEF1A2 ( $P=0.01$ ). To extend these findings, we then studied 11 tissue microarray blocks comprising normal and tumor tissue specimens from 113 patients with pathologic stage I non-small-cell lung cancer who had undergone curative surgery (Wang *et al.*, 2005). Immunohistochemical analysis showed that EEF1A2 was expressed in 32 cases (28%) and KCIP-1 in 29 cases (26%). Univariate and multivariate Cox proportional hazards models were used to detect possible associations between EEF1A2 and KCIP-1 expression and clinicopathologic variables. Expression of EEF1A2 or KCIP-1 was associated with short overall survival time ( $P=0.0012$  for EEF1A2 and  $P=0.0026$  for KCIP-1) (Supplementary Figure 6S). Age at diagnosis, histologic type of cancer, degree of tumor differentiation, and smoking history were not associated with survival time.

Although only two genes were validated in the lung tissue microarrays (because available antibodies to the other two genes were not suitable for use in

**Table 2** Proteins showing significant overexpression in cancer cell lines relative to those in normal bronchial epithelial cell lines and their correlation coefficients with increased DNA copy number or mRNA values\*

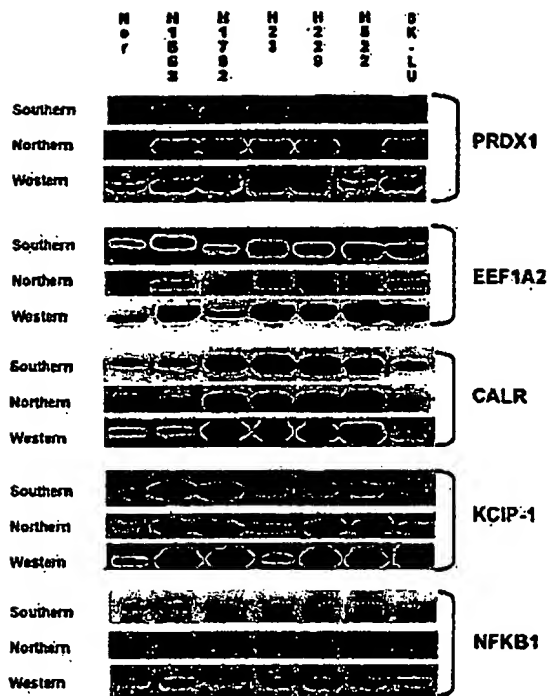
Acc. no.	Gene ID	Gene	Mw/pI	Description	r with genomic copy changes <sup>a</sup>	r with mRNA changes <sup>b</sup>
Q06830	5052	PRDX1	48.4/5.4	Peroxiredoxin 1	0.92364	0.91892
Q05639	1917	EEF1A2	50.5/5.7	Eukaryotic translation elongation factor 1 alpha 2	0.90218	0.89456
P27797	811	CALR	61/5.5	Calreticulin	0.84128	0.86434
P63104	7534	KCIP-1	27/6.5	Tyrosine 3-monooxygenase activation protein, zeta	0.84467	0.85499
P07237	5034	P4HB	54/6.2	Procollagen-proline, 2-oxoglutarate 4-dioxygenase	0.91884	0.76786
Q04695	3872	KRT17	48.0/4.9	Keratin 17	0.00236	0.86892
P09211	2950	GSTP1	23.2/4.7	Glutathione S-transferase pi	0.84218	0.69456
P17936	3486	IGFBP-3	31.6/5.8	Insulin-like growth-factor binding protein 3	0.06412	0.16434
P26641	1937	EEF1G	50/6.4	Eukaryotic translation elongation factor 1 gamma	0.00446	0.85549
P08727	3880	KRT19	44.1/5.2	Keratin 19	-0.04884	0.86786
P04792	3315	HSPB1	22/6.5	Heat shock 27kDa protein 1	0.00364	0.31892
P00558	5230	PGK1	44.5/4.2	Phosphoglycerate kinase 1	0.50402	0.79456
Q01995	5876	TAGLN	22.5/4.3	Transgelin	-0.34128	-0.26434
P08631	3055	JTK9	59.5/6.8	Hemopoietic cell kinase	-0.01446	0.02549
P09382	3956	LGALS1	16/5.5	Galectin-1, galactoside-binding, soluble, 1	0.026623	0.01123
Q92784	8110	DPF3	25.8/4.8	D4, zinc and double PHD fingers, family 3	0.094884	-0.03214
P54257	9001	HAP1	75.5/6.5	Huntington-associated protein 1	0.12364	-0.08108
P05783	3875	KRT18	48/5.3	Keratin 18	0.010218	0.60544
P05787	3856	KRT8	9.2/4.4	Keratin 8	0.041280	0.84566
P00738	3240	HP	55.2/6.2	Haptoglobin	0.044679	-0.14501
P09769	2268	FGR	59.5/5.2	Gardner-Rasheed feline sarcoma viral oncogene homolog	0.031264	-0.13789
P19838	4790	NFKB1	50.4/6.3	Nuclear factor of kappa light gene enhancer in B-cells 1	0.04467	-0.14501
P29034	6273	S100A2	10.9/4.6	S100 calcium-binding protein A2	0.87964	0.243214
Q13105	7709	ZBTB17	87.9/5.3	Zinc-finger and BTB domain containing 17	-0.17636	0.048108
Q00987	4193	MDM2	75.2/4.8	Transformed 3T3 cell double minute 2	-0.19782	-0.50544
P27816	4134	MAP4	111/5.4	Microtubule-associated protein 4	0.25872	-0.05356
P52732	3832	KIF11	119.2/6.2	Kinesin family member 11	-0.25778	-0.53444
P25205	4172	MCM3	90.9/5.5	Minichromosome maintenance deficient 3	0.25644	0.053666
P08631	3055	HCK	59.5/5.7	Hemopoietic cell kinase	0.65533	0.054501
P09237	4316	MMP7	22.6/5.8	Matrix metalloproteinase 7	0.234987	0.876820
P30305	994	CDC25B	64.9/4.5	Cell division cycle 25B	0.045116	0.283214
P50290	998	CDC42	21.3/6.1	Cell division cycle 42 (GTP-binding protein, 25kDa)	-0.47636	0.088108
P61586	387	RHOA	19.8/6.9	Ras homolog gene family, member A	-0.49782	-0.00544
P63000	5879	RAC1	21.5/6.8	Ras-related C3 botulinum toxin substrate 1	-0.05583	-0.03566
P07437	203068	TUBB	49.6/6.5	Tubulin, beta polypeptide	0.255533	0.145010
P24864	898	CCNE1	47.1/4.3	Cyclin E1	-0.65116	0.232149
P04141	1437	CSF2	16.9/6.3	Colony stimulating factor 2 (granulocyte-macrophage)	-0.64636	-1.28108
P28072	5694	PSMB6	25.3/5.2	Proteasome (prosome, macropain) subunit, beta type, 6	-0.69782	-1.30544
P00352	216	ALDH1A1	54.7/4.3	Aldehyde dehydrogenase 1 family, member A1	-0.75872	0.03356
Q03013	2948	GTM4	25.3/5.0	Glutathione S-transferase M4	-0.78533	0.134501
P63241	1984	EIF5A	10/4.4	Eukaryotic translation initiation factor 5A	-0.97893	-1.44321
Q01469	2171	EFABP	18.0/4.2	Fatty acid-binding protein 5	0.25684	-0.36432

\*Only the gene showing statistically significant increased protein expression with increases in both genomic copy number and transcript simultaneously will be considered as potential putative oncogene in lung adenocarcinoma cells. <sup>a</sup>r, Spearman correlation coefficients between proteins and genomic or mRNA values are based on all six cancer cell lines; bold indicates  $P < 0.05$ , if  $r > 0.84000$ . Mw, molecular weight; pI, isoelectric point.

immunohistochemical analysis), these findings are consistent with those from our cell lines, demonstrating again that genomic amplification and consequent increases in amounts of transcript may be, at least in part, driving the abundance of proteins in these lung tumors. The association between expression of these genes and that of Ki-67, a known indicator of poor prognosis in lung cancer (Martin *et al.*, 2004), suggests that activation of these genes may be an indicator of tumor aggressiveness. These results also suggest that expression of EEF1A2 and KCIP-1 proteins in stage I non-small-cell lung cancer may be useful as a marker for distinguishing patients with relatively poor prognosis from those who might benefit from adjuvant treatment.

## Discussion

Our current study illustrates the power of integrated functional genomic analyses for identifying putative oncogenes and for evaluating their potential clinical significance. Among the four identified oncogenes, three genes (*PRDX1*, *CALR*, and *KCIP-1*) have been implicated in lung tumorigenesis. *PRDX1* is an antioxidant protein involved in regulating cell proliferation, differentiation, and apoptosis. Kim *et al.* (2003) found *PRDX1* expression to be elevated in both lung cancer and adjacent normal lung tissue, suggesting that activation of *PRDX1* may enhance proliferation in lung cancer. *CALR* has a major role in  $Ca^{2+}$  binding and the



**Figure 1** Confirmation by Southern, Northern, and Western blot analyses of increased DNA copies, transcript levels, and protein levels in the four genes identified in high-throughput analyses. For comparison, we arbitrarily chose one gene, *NFKB1*, in which an increased protein level did not correlate with genetic changes. The blotting results are consistent with the results from the CGH array, transcript array, and proteomic analyses. Nor, indicates normal bronchial epithelial cell line. All the experiments were repeated at least three times with each cell line. Means of normalized to  $\beta$ -actin signal intensities on Southern, Northern, and Western blots, along with 95% confidence intervals, were calculated ( $\beta$ -actin signals are not shown in the figure; two different normal bronchial epithelial cell lines were used in the confirmation and only one normal cell line is shown in the figure).

transcriptional regulation of other genes and was recently found to be overexpressed in 73% of 40 lung adenocarcinomas (Oates and Edwards, 2000). *KCIP-1* belongs to the 14-3-3 family, which participates via the MAPK and Wnt signaling pathways in the regulation of many cellular processes including cell proliferation and differentiation as well as tumorigenesis (Thomas et al., 2005). *KCIP-1* was recently found to be expressed in all 12 lung tumors tested in a single-institution study (Qi et al., 2005). Interestingly, *EEF1A2* was originally considered a putative oncogene in ovarian cancer on the basis of its being amplified in 25% and overexpressed in 30% of the same set of ovarian tumors (Anand et al., 2002); functional analyses have established its oncogenic role in cellular transformation (Lee, 2003). Our discovery that *EEF1A2* may be a putative oncogene in lung adenocarcinoma demonstrates the power of our functional genomic strategy for rapidly identifying potential oncogenes.

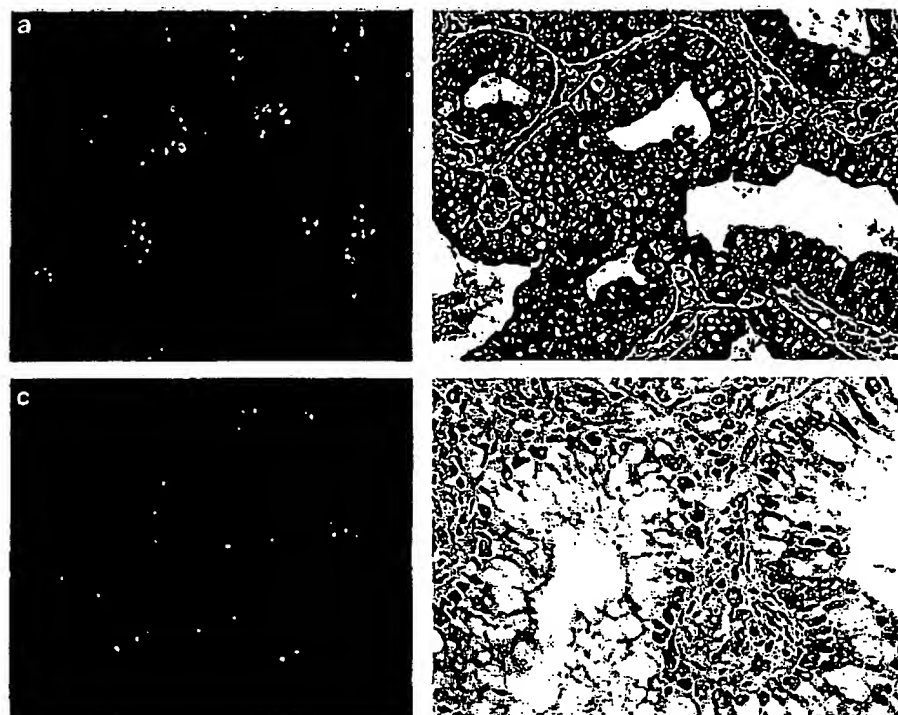
Although the main focus of this study was to specifically identify putative oncogenes, it should be

noted that 90.7% of the genes showing high protein expression did not show corresponding increases in both DNA copy number and transcript, a finding consistent with that of others that transcriptional, translational, and post-translational regulatory mechanisms can greatly influence the abundance of protein in lung tumorigenesis (Chen et al., 2002). For example, *NFKB1* is a critical arbiter of immune responses, cell survival, and transformation and is often activated in several types of tumors (Chen et al., 2002). Deregulation of *NFKB1* is thought to be modulated through phosphorylation of Ser337 by protein kinase A (Chen et al., 2002). In our study, 68.8% of the genes showing over-representation in the genome did not show elevated transcript levels, implying that at least some of these genes are 'passenger' genes that are concurrently amplified because of their location with respect to amplicons but lack biological relevance in terms of the development of lung adenocarcinoma.

Although the potential oncogenes we identified here are likely to be important, certainly other oncogenes could be involved in the development of lung adenocarcinoma. The oligo microarray we used consists of 22 000 probes, which represent only about 60% of the human genome. Moreover, each probe was designed for the 3' region of expressed sequence tags of the selected genes. Also, our results were initially derived from cancer cell lines, although the findings were later confirmed in human tissue samples. Our ongoing study using microarrays with information on more genes and the development of high-resolution proteomic analyses for use with larger numbers of specimens will allow more comprehensive analyses of the molecular consequences of gene amplifications. Such expanded analyses will very likely lead to the identification of additional oncogenes.

Some of the results of our current study were comparable to those of other studies of lung cancer. For example, genomic copy number and protein levels of *KCIP-1* were previously found to be amplified and overexpressed in primary lung cancers by cDNA clone-based CGH array analysis (Jiang et al., 2004) and proteomic analysis (Chen et al., 2002), respectively. Our functional genomic approach, which integrates simultaneous CGH, transcript microarrays, proteomic analyses, and siRNA, allows us not only to quickly identify potential oncogenes but also to explore their significance as diagnostic and therapeutic targets in tumor progression – more than could be achieved by any technique alone.

Genes identified in this way may serve as promising targets for diagnosis and therapy in lung adenocarcinoma. Further research on the clinical implications of such genes is needed; experiments now underway in our laboratory include overexpression of the genes in normal cells, disruption of the function of these genes in cancer cells, and investigation of how interactions among these genes (or interactions with other known oncogenes) may mediate the expression of the transformed phenotype.



**Figure 2** *EEF1A2* amplification is associated with high *EEF1A2* protein expression in lung adenocarcinomas. (a) Cells from a lung adenocarcinoma sample in which *EEF1A2* is amplified show more green signals (*EEF1A2*) than red signals (chromosome 20 centromeric probe) (original magnification,  $\times 400$ ). (b) Immunohistochemical staining of cells from the same tissue sample as in panel a shows strong *EEF1A2* staining in the cytoplasm. (c) A lung adenocarcinoma sample with two copies of *EEF1A2* and chromosome 20 centromeric probe, indicating no *EEF1A2* amplification (original magnification,  $\times 400$ ). (d) Immunohistochemical staining of cells from the same tissue sample as in panel c shows negative staining for *EEF1A2*.

## Materials and methods

### Cell lines

Six human lung adenocarcinoma cell lines (H23, H229, H1792, SK-LU-1, H522, and H1563) were obtained from the American Type Culture Collection (Manassas, VA, USA). Two normal bronchial epithelial cell lines were obtained from Clontech (Palo Alto, CA, USA). Genomic DNA, mRNA, and protein were derived from a single harvest of these cells.

### DNA and RNA profiles by microarray analysis

Genomic DNA labeling and hybridization were performed as described previously (Barrett *et al.*, 2004) with Agilent's Human 1A Oligo Microarray (V2) (Agilent Technologies, Palo Alto, CA, USA), which contains 22000 unique 60-mer oligos. Details of the protocol for analysing transcripts are available at <http://www.chem.agilent.com>. Map positions for arrayed genes were assigned by identifying the DNA sequence represented in the UniGene cluster and matching it with the Golden Path genome assembly (<http://genome.ucsc.edu/>; Mat 7, 2004 Freeze). Microarray images of DNA copy number and expression were analysed by using AgilentCGH Analytics and Feature Extraction software. DNA copy number profiles that deviated significantly from background signal ratios (measured from normal control cell hybridization, as described elsewhere; Barrett *et al.*, 2004) were interpreted as evidence of true differences in DNA copy number. The criteria for defining genomic over-representation and amplicons are described elsewhere (Hyman *et al.*, 2002); details are given in the

Supplementary Information. An increase in mRNA level was defined as a twofold increase in signal ratio relative to that of the control ( $\log_2 > 1$ ).

### Quantitative two-dimensional PAGE and mass spectrometry

Analysis of proteins by two-dimensional PAGE and their identification by mass spectrometry were performed as previously described (Shen *et al.*, 2004). Briefly, protein pellets were solubilized in rehydration buffer, after which the first-dimension isoelectric focusing was carried out with a Protean IEF Cell (Bio-Rad Laboratories) and the second-dimension separation was carried out with Bio-Rad's Ready Gel Precast Gels and the Bio-Rad Criterion Cell apparatus. Protein spots were visualized by silver-based staining, and all gels were assessed with Bio-Rad's PDQuest 2D gel image analysis software. Selected spots were subjected to in-gel tryptic digestion and analysed on a Voyager-DE PRO matrix-assisted laser desorption ionization/time-of-flight mass spectrometer (Applied Biosystems, Foster City, CA, USA). The mass list of the 20 most intense monoisotopic peaks for each sample was entered in the MS-Fit search program (v3.2.1) (<http://prospector.ucsf.edu/ucsfhtml4.0/msfit.htm>) and searched in the National Center for Biotechnology Information protein database.

### Southern, Northern, and Western blot analyses

Southern, Northern, and Western blot hybridizations were performed according to standard protocols. cDNA clones for the tested genes were purchased from Invitrogen (Carlsbad,

CA, USA) and prepared as probes for the blot hybridizations. Antibodies used were obtained as follows: PRDX1, CALR, NFKB1, KCIP-1, and  $\beta$ -actin from Santa Cruz Biotechnology (Santa Cruz, CA, USA); and EEFA2 from Upstate Biotechnology (Waltham, MA, USA).

#### Fluorescence in situ hybridization and immunohistochemical analyses of lung tissue microarrays

Fluorescence in situ hybridizations and immunohistochemical analyses of KCIP-1 and EEFA2 were carried out as described elsewhere (Jiang et al., 2002; Wang et al., 2005) with Lung Tissue Microarrays (Ambion, Austin, TX, USA) and 11 homemade microarray blocks containing tissue samples from 113 patients with pathologic stage I non-small-cell lung cancer (Wang et al., 2005). DNA probes specific for KCIP-1 and EEFA2 were obtained by screening a Human BAC Clone library (Invitrogen) by polymerase chain reaction as described previously (Jiang et al., 2002). The antibodies used for the immunohistochemical analyses were the same as those used for the Western blotting. Cell proliferation of the lung tissues was assessed with a Ki-67 monoclonal antibody from Santa Cruz Biotechnology. Definitions of the cutoff value for a positive result of each antibody are shown in Supplementary Information.

#### siRNA transfection, cellular proliferation assay, and apoptosis analysis

Transfections were carried out by using siPORT Lipid Transfection Agent (Ambion) with siRNAs targeting KCIP-1 or EEFA2 or with a scrambled siRNA duplex (siControl) (Dharmacon Inc., Lafayette, CO, USA), with PBS used as a negative control (Jiang et al., 2002). Cells were fixed 24, 48, or 96 h later and subjected to further tests. All siRNAs were prepared by using a transcription-based method with Silencer siRNA according to the manufacturer's instructions (Ambion). Sequences of the individual siRNAs are listed in Supplementary Table 4S. Inhibition of cell growth by the

siRNAs was determined by MTT staining, and cell growth rate was plotted against the percentage of viable cells in the saline-treated controls (a value arbitrarily set at 100%) (Jiang et al., 2002). Apoptosis was analysed by fluorescence cell cycle analysis of terminal deoxynucleotidyl transferase-mediated dUTP nick-end labeling with FITC-labeled dUTP (Boehringer Mannheim Biochemicals, Mannheim, Germany) (Jiang et al., 2005).

#### Statistical analyses

Relationships between gene copy number and mRNA level were examined as described elsewhere (Hyman et al., 2002, Supplementary Information). Correlations between protein abundance and DNA copy number and mRNA expression of the corresponding genes were evaluated with the Spearman correlation coefficient. Fisher's exact test and  $\chi^2$ -tests were used to analyse associations between amplification and expression of the candidate genes with various histopathologic variables of the samples in the tissue microarrays. Univariate and multivariate analyses were carried out with Cox's proportional hazards model to determine which independent factors might have a joint significant influence on survival. A *P*-value  $\leq 0.05$  was considered statistically significant; all statistical tests were based on a two-sided significance level.

#### Acknowledgements

This work was supported by National Institutes of Health Grant CA113707-01, an institutional research grant from The University of Texas MD Anderson Cancer Center, a Developmental Project/Career Development Award from The University of Texas Specialized Programs of Research Excellence in Lung Cancer P50 CA70907, and an M Keck Center for Cancer Gene Therapy Award (all to FJ). We thank Christine F Wogan of the Department of Scientific Publications for editorial review of this manuscript.

#### References

- Anand N, Murthy S, Amann G, Wernick M, Porter LA, Cukier IH et al. (2002). *Nat Genet* 31: 301–305.
- Barrett MT, Scheffer A, Ben-Dor A, Sampas N, Lipson D, Kincaid R et al. (2004). *Proc Natl Acad Sci USA* 101: 17765–17770.
- Chen G, Gharib TG, Huang CC, Taylor JM, Misk DE, Kardias SL et al. (2002). *Mol Cell Proteomics* 1: 304–313.
- Hyman E, Kauraniemi P, Hautaniemi S, Wolf M, Mousset S, Rozenblum E et al. (2002). *Cancer Res* 62: 6240–6245.
- Jiang F, Caraway NP, Li RY, Katz RL. (2005). *Oncogene* 24: 3409–3418.
- Jiang F, Lin F, Price R, Gu J, Medeiros LJ, Zhang HZ et al. (2002). *J Mol Diagn* 4: 144–149.
- Jiang F, Yin Z, Caraway NP, Li R, Katz RL. (2004). *Neoplasia (New York)* 6: 623–635.
- Kim HJ, Chae HZ, Kim YJ, Kim YH, Hwang TS, Park EM et al. (2003). *Cell Biol Toxicol* 19: 285–298.
- Lee JM. (2003). *Reprod Biol Endocrinol* 1: 69–73.
- Luk C, Tsao MS, Bayani J, Shepherd F, Squire JA. (2001). *Cancer Genet Cytogenet* 125: 87–99.
- Martin B, Paesmans M, Mascaux C, Berghmans T, Lothaire P, Meert AP et al. (2004). *Br J Cancer* 91: 2018–2225.
- Oates J, Edwards C. (2000). *Histopathology* 36: 341–347.
- Pollack JR, Sorlie T, Perou CM, Rees CA, Jeffrey SS, Lonning PE et al. (2002). *Proc Natl Acad Sci USA* 99: 12963–12968.
- Qi W, Liu X, Qiao D, Martinez JD. (2005). *Int J Cancer* 113: 359–363.
- Ross JS, Fletcher JA. (1999). *Semin Cancer Biol* 9: 125–138.
- Shen J, Person MD, Zhu J, Abbuzzese JL, Li D. (2004). *Cancer Res* 64: 9018–9026.
- Thomas D, Guthridge M, Woodcock J, Lopez A. (2005). *Curr Top Dev Biol* 67: 285–303.
- Tonon G, Wong KK, Maulik G, Brennan C, Feng B, Zhang Y et al. (2005). *Proc Natl Acad Sci USA* 102: 9625–9630.
- Wang H, Zhang Z, Li R, Ang KK, Zhang H, Caraway NP et al. (2005). *Int J Cancer* 116: 285–290.

Supplementary Information accompanies the paper on Oncogene website (<http://www.nature.com/onc>)

# Genetic Instability in Epithelial Tissues at Risk for Cancer

WALTER N. HITTELMAN

*Department of Experimental Therapeutics, The University of Texas  
M. D. Anderson Cancer Center, Houston, Texas 77030, USA*

**ABSTRACT:** Epithelial tumors develop through a multistep process driven by genomic instability frequently associated with etiologic agents such as prolonged tobacco smoke exposure or human papilloma virus (HPV) infection. The purpose of the studies reported here was to examine the nature of genomic instability in epithelial tissues at cancer risk in order to identify tissue genetic biomarkers that might be used to assess an individual's cancer risk and response to chemopreventive intervention. As part of several chemoprevention trials, biopsies were obtained from risk tissues (i.e., bronchial biopsies from chronic smokers, oral or laryngeal biopsies from individuals with premalignancy) and examined for chromosome instability using *in situ* hybridization. Nearly all biopsy specimens show evidence for chromosome instability throughout the exposed tissue. Increased chromosome instability was observed with histologic progression in the normal to tumor transition of head and neck squamous cell carcinomas. Chromosome instability was also seen in premalignant head and neck lesions, and high levels were associated with subsequent tumor development. In bronchial biopsies of current smokers, the level of ongoing chromosome instability correlated with smoking intensity (e.g., packs/day), whereas the chromosome index (average number of chromosome copies per cell) correlated with cumulative tobacco exposure (i.e., pack-years). Spatial chromosome analyses of the epithelium demonstrated multifocal clonal outgrowths. In former smokers, random chromosome instability was reduced; however, clonal populations appeared to persist for many years, perhaps accounting for continued lung cancer risk following smoking cessation.

**KEYWORDS:** chromosome instability; epithelial cells; aerodigestive tract; chemoprevention; cancer risk

## THE NEED FOR BIOMARKERS OF CANCER RISK AND RESPONSE TO INTERVENTION

Epithelial cancers remain a major health challenge in the world. Despite improvements in staging and the application and integration of surgery, radiotherapy, and chemotherapy, the 5-year survival rate for individuals with lung cancer is only about 15%.<sup>1</sup> Even if strategies for early detection are successful and lung cancers are detected at a stage where local tumor resection and treatment is curative, these patients will still be at significant risk for developing second primary tumors

Address for correspondence: Dr. Walter N. Hittelman, Department of Experimental Therapeutics, The University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Blvd. (Box 19), Houston, Texas 77030. Voice: 713-792-2961; fax: 713-792-3754.  
whittelm@mdanderson.org

associated with the problem of field cancerization.<sup>2</sup> Similarly, for individuals with a first head and neck primary tumor, even if the first malignancy is successfully treated, the risk of developing a second primary in the tobacco smoke-exposed field is approximately 40%.<sup>3</sup> Similar cancer risk estimates exist for individuals who exhibit severe dysplasia in premalignant epithelial lesions.<sup>4</sup> For these reasons, it is important to focus on chemopreventive strategies to prevent the development of epithelial malignancies.

Several problems confront chemoprevention trials designed to identify efficacious agents.<sup>5</sup> First, chemoprevention trials with cancer incidence as a primary endpoint require tens of thousands of subjects and tens of years of intervention and follow-up for statistical evaluation. For example, a recently reported trial involved 30,000 subjects and required 10 years in order to examine the impact of prevention strategies on lung cancer development, only to find a possible increased lung cancer incidence in current smokers who received  $\beta$ -carotene.<sup>6</sup>

The problem of large, long-term trials results from the difficulty in identifying individuals at highest cancer risk who might best benefit from chemopreventive intervention. For example, 20 pack-year smokers, while known to be at relatively increased risk for developing lung cancer, have approximately a 10% lifetime risk for developing lung cancer.<sup>7</sup> This seriously limits the number of potentially useful strategies that can be clinically explored. A second problem facing chemoprevention trials is that little is known about what agents are likely to have efficacy, and even less is known regarding proper doses, schedules, and durations of treatment. Part of the reason for this problem is that too little is known about the physiologic processes that drive epithelial cancer development.

In order to reduce the number of subjects and the time required to carry out chemoprevention trials and thus allow the exploration of multiple prevention strategies, two types of advances are necessary. First, it is important to identify individuals at significantly increased cancer risk who might best benefit from different types of intervention. Second, in order to allow the rapid identification of agents, doses, and schedules of potentially efficacious agents, it is necessary to identify and validate surrogate endpoints of response that indicate whether the agents are having a positive impact on the target tissue during the chemopreventive intervention.

One approach to identifying individuals at increased aerodigestive tract cancer risk is to explore epidemiologic features of potential subjects. Molecular epidemiologic studies are beginning to identify intrinsic host factors that place some individuals at increased cancer risk, especially those with a chronic smoking history.<sup>8</sup> Most intrinsic factors identified thus far reflect levels of carcinogen metabolism, repair capabilities of the host following DNA damage, and other measures of intrinsic cellular sensitivity to mutagens. While these factors can provide statistically significant risk ratios in case-control studies that are controlled for tobacco exposure, the detected risk ratios usually fall in the range of 1.5 to 10. Unfortunately, this is not sufficient for the individualization of treatment and is not sufficiently high to significantly reduce the numbers of subjects required for chemoprevention trials with cancer incidence as the primary endpoint.

Another approach to identifying individuals at increased cancer risk is to directly examine the target tissue of individuals with known carcinogen exposure (e.g., chronic tobacco smoke exposure), who have evidence of target organ dysfunction

(e.g., chronic obstructive pulmonary disease, changes in voice quality), or who have clinical evidence of premalignancy (e.g., bronchial metaplasia/dysplasia, oral leukoplakia/erythroplakia, cervical intraepithelial neoplasia). The conventional standard for assessing cancer risk in these situations is the degree of histological change. However, while individuals who show moderate to severe dysplasia are known to be at increased cancer risk when compared to individuals with lesser histologic changes, it is often difficult to distinguish reactive changes to carcinogenic insult from initiated and progressing lesions. Similarly, upon cessation of carcinogenic insult, histologic changes may reverse yet cancer risk may continue for many years. For example, while smoking cessation is associated with decreased bronchial metaplasia,<sup>9</sup> increased lung cancer risk continues for many years beyond smoking cessation.<sup>10</sup> In fact, nearly half the newly diagnosed lung cancer cases in the USA occur in former smokers.<sup>11</sup>

The development of assays to identify individuals at high epithelial cancer risk and to directly assess response to intervention in the target tissue is therefore an important research goal. Such assays should be objective and easily quantifiable and, if possible, minimally invasive. Moreover, they should reflect both the disease process and the targeted pathway and thereby be useful in assessing risk and monitoring response to intervention as well as directly testing the hypothesized mechanism of action of the chemopreventive strategy.

In the chemoprevention setting it is important to recognize that one does not know the location of the future cancer. Thus, assays must necessarily be carried out on random biopsies of the field at risk. Even if there are clinically evident premalignant lesions, this does not mean that this is the likely site for a future malignancy. For example, nearly half of the cancers that develop in individuals with oral leukoplakia arise away from the original index lesion. Similarly, since many newly diagnosed lung cancers arise in the peripheral parts of the lung (e.g., adenocarcinomas), especially in former smokers, and since endobronchoscopy predominantly accesses central components of the lung, it is important to identify biomarkers that can reflect global processes ongoing in the target epithelial field associated with increased cancer risk. Their discovery requires a better understanding of the tumorigenesis process in epithelial fields at cancer risk.

#### THE RATIONALE FOR STUDYING GENOMIC INSTABILITY AS A MARKER OF RISK

Tumors of the aerodigestive tract have been proposed to reflect a "field cancerization" process whereby the whole tissue is exposed to carcinogenic insult (e.g., tobacco smoke) and is at increased risk for multistep tumor development.<sup>12,13</sup> Several types of clinical and laboratory data support this notion, including the frequent occurrence of synchronous primary and subsequent second primary tumors in the aerodigestive tract (frequently exhibiting dissimilar histologies as well as distinct genetic signatures<sup>14-16</sup>) and the presence of premalignant lesions that precede and/or accompany the tumor in the exposed tissue field.<sup>17</sup> The notion of a multistep tumorigenesis process is further supported by serial clinical and histologic evaluations of

target tissue or exfoliated cells where increasing degrees of histological abnormalities are observed over time.<sup>18</sup>

A working model for aerodigestive tract tumorigenesis is illustrated in FIGURE 1. Tumorigenesis in the face of carcinogenic exposure likely involves a chronic process of tissue injury and wound healing. DNA damage induced by the carcinogen is likely fixed into permanent genetic changes (e.g., chromosome damage, chromosome non-disjunction, gene mutation, gene deletion, etc.) during the process of proliferation. This damage would be expected to be distributed throughout the exposed tissue field leading to a background of generalized genomic damage (depicted in FIGURE 1 as a background mat of increasing density). Chronic injury and repair likely leads to the accumulation of cells with increasing amounts of genetic changes as well as the outgrowth of abnormal clones (triangles in FIGURE 1) carrying an accumulation of genetic changes important for selective survival, dysregulated growth, and preferential epithelial take-over by initiated clones (see FIGURE 2).

Cellular and molecular evidence for the field carcinogenesis and multistep tumorigenesis model comes from many laboratories.<sup>19,20</sup> With the advent of a wide array of molecular technologies, a large number of specific molecular genetic and epigenetic changes involving specific oncogenes, tumor suppressor genes, cell regulatory genes, and repair genes have now been described for aerodigestive tract cancers. The identification of these specific molecular changes have now provided probes to explore specific events occurring in premalignant lesions adjacent to aerodigestive tract tumors.<sup>21-24</sup> Frequently, these premalignant lesions showed a subset of the same molecular changes found in the associated tumor, suggesting that these lesions might represent precursor lesions for the associated tumors (i.e., a manifestation of

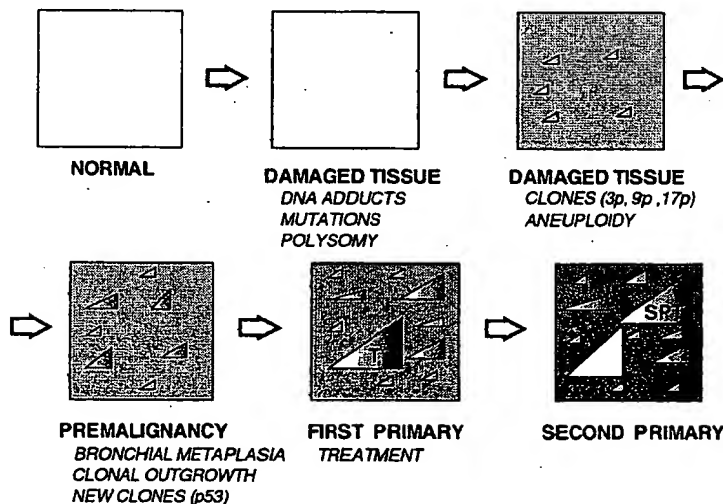


FIGURE 1. Field cancerization and multistep tumorigenesis.

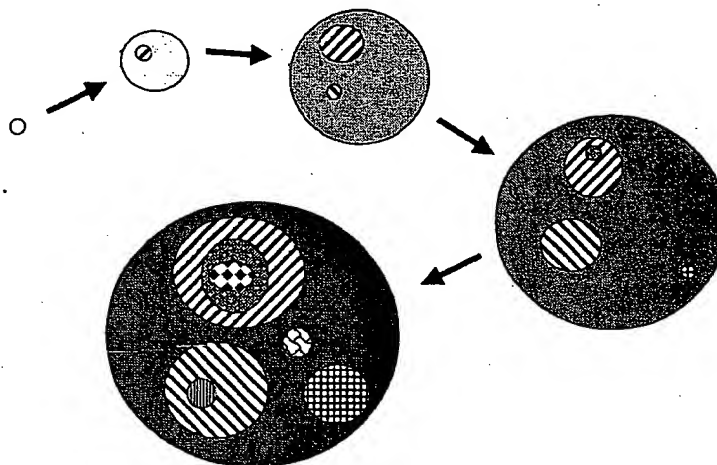


FIGURE 2. Multiple focal clonal evolution during multistep tumorigenesis.

a multistep tumorigenesis process). For example, studies of the premalignant lesions adjacent to head and neck tumors have provided evidence for a gradual accumulation of genetic alterations accompanied by evidence for dysregulation of cellular control mechanisms (e.g., alterations in expression of PCNA, EGFR, TGF- $\beta$ , p53, and cyclin D1).<sup>25-28</sup>

These types of studies have now also been applied to the target epithelium of individuals at increased risk for aerodigestive tract cancer (i.e., individuals with a chronic smoking/alcohol history and/or prior aerodigestive tract cancer). Several groups (using polymerase chain reaction, PCR, analysis of microdissected epithelium) have now demonstrated the presence of clonal outgrowths in the target premalignant epithelium of individuals at increased risk for cancer.<sup>29-31</sup> For example, examination of bronchial biopsies derived from individuals with a 20 pack-year smoking history demonstrated that 76% of the cases showed evidence for LOH (3p14, 9p21, or 17p13) in at least one of six lung biopsy sites. On a per site basis, some form of LOH was observed in 25% of the sites examined.<sup>29</sup>

If aerodigestive tract cancer development reflects a field cancerization process involving multistep events, then risk and response information should be able to be derived from random biopsies or exfoliated cells from the field at risk or from assessments of tissue undergoing similar processes. Hypothetically, lesions exhibiting the greatest degree of genomic instability, clonal outgrowth, and abnormal epithelial regulation would be at the highest relative aerodigestive tract cancer risk. Similarly, an active chemopreventive intervention might be expected to decrease these manifestations of risk. Reduced risk manifestations include decreased levels of ongoing genetic instability, decreased frequency of clonal outgrowths, and increased epithelial growth regulation.

### THE MEASUREMENT OF CHROMOSOME INSTABILITY USING CHROMOSOME *IN SITU* HYBRIDIZATION

Molecular genetic techniques, while extremely useful for detecting clonal changes in target tissues, are somewhat limited in their ability to detect random genetic instability. Conventional cytogenetic assays are useful for detecting chromosome instability and clonal chromosome changes. However, they require numbers of dividing cells for karyotypic analysis that are difficult to attain in the setting of biopsies acquired during the course of a chemoprevention trial. A technique was therefore needed that would allow chromosome instability measurements in situations where few cells are available (e.g. small biopsies, brushings, or sputum samples) and where the target material might be fixed. It was also desirable to have a technique that would be adaptable to tissue sections, whereby spatial information could be retained and genotype/phenotype associations could be determined on the same or adjacent tissue sections. The technique of *in situ* hybridization (ISH) involves the use of DNA probes that recognize either chromosome-specific repetitive target sequences, chromosome single gene copy sequences, or sequences along the whole chromosome length or chromosome segments.<sup>32</sup> We have adapted the ISH technique for formalin-fixed, paraffin-embedded tissue sections and have applied it to a variety of tissues, including the aerodigestive tract.<sup>33,34</sup>

Using probes that label the centromere regions of specific chromosomes, this assay permits determination of the average chromosome number per cell for each specimen. This assay is also useful for detecting generalized chromosome instability during the tumorigenesis process. Normal diploid populations should have two copies of each autosomal chromosome and should rarely show three or more chromosome copies per cell (chromosome polysomy), especially in tissue sections where nuclear truncation results in an under-representation of chromosome copy number. Thus, the detection of cells with three or more chromosome copies would indicate the presence of chromosome instability.

To examine this technique's potential for characterizing the multistep tumorigenesis process in the aerodigestive tract, we measured the fraction of cells exhibiting three or more chromosome copies in apparently contiguous epithelial transitions from normal to hyperplastic to dysplastic to carcinomas, all on a single tissue slice of head and neck squamous cell carcinomas.<sup>34</sup> In these specimens, greater than 35% of the cases of adjacent "normal" epithelium, greater than 65% of the cases of hyperplastic epithelium, and greater than 95% of the dysplastic and tumor regions showed evidence of chromosome polysomy. Of interest, similar transitions of chromosome instability were observed with at least four different chromosome probes. Similar trends have also been observed in amenable tissue from other epithelial malignancies, including cervix, bladder, and breast.<sup>35</sup> These results thus suggested that the notions of field cancerization and multistep tumorigenesis might apply to several epithelial tissues and that measures of chromosome instability might be useful for monitoring this process.

In the situations described above, the premalignant lesions examined might be considered to represent epithelium at 100% risk of being in a cancer field, since they were located in the adjacent epithelium to the cancer. This then raises the question of the nature of genetic instability in the epithelium of individuals at increased risk

for developing cancer. To explore this issue, we obtained biopsies during the course of leukoplakia chemoprevention trials exploring the use of 13-*cis*-retinoic acid in reversing leukoplakia and probed them for genetic instability using *in situ* hybridization. In one retrospective study and in one prospective study of subjects with oral leukoplakia, the results indicate that those subjects whose pretreatment biopsies harbor relatively high levels of genomic instability (i.e., more than 3% of the cells examined showing at least 3 chromosome 9 copies per cell) have a significantly higher likelihood of suffering early onset of head and neck cancer.<sup>36,37</sup> Interestingly, half of the tumors that did develop occurred away from the biopsy site used to measure genetic instability. This result suggests that genomic instability measurements in carcinogen-exposed tissue can provide useful cancer risk estimates.

### THE RELATIONSHIP BETWEEN TOBACCO EXPOSURE AND CHROMOSOME INSTABILITY

In recent years, the aerodigestive tract chemoprevention group at M.D. Anderson Cancer Center has initiated three sequential biomarker-associated chemoprevention trials involving chronic smokers with a greater than 20 pack-year smoking history. In each of these studies, endobronchial biopsies were obtained from six defined sites within the lung, including the carina and at bifurcation points at the upper, middle, and lower right lung and at the upper and lower left lung. Biopsies were obtained prior to and following chemopreventive intervention and were subjected to *in situ* hybridization analysis in addition to analyses for other biomarkers. The first important finding was that some degree of chromosome polysomy was evident in all lung sites examined, and this was observed independently of the particular chromosome probe utilized.<sup>38</sup> This finding supports the notion that random chromosome changes may be occurring throughout the exposed lung field.

In a second study, bronchial biopsies were obtained from individuals with a 20 pack-year smoking history. In this study, most of the subjects involved were current smokers.<sup>39</sup> Interestingly, all cases who showed metaplasia at one of six biopsy sites also showed chromosome polysomy in at least one biopsy site; overall, 88% of the sites showed some evidence of chromosome 9 polysomy.<sup>40</sup> Evidence for genetic instability was also detected in patients who did not show evidence of bronchial metaplasia in any of six biopsy sites despite a strong smoking history. In fact, more than 90% of the cases and more than 60% of the sites showed significant chromosome polysomy (i.e., at least three copies in at least 2 % of the cells examined). These results suggest that the lungs of long-term smokers show significant evidence of genetic instability, and this instability can be detected throughout the accessible bronchial tree, even when bronchial metaplasia is not evident.

These studies in current smokers has allowed us to examine the relationship between the levels of genetic instability detected and subject characteristics such as smoking status (current or former), smoking history, and lung tissue pathologic changes. Evaluable biopsy material has now been obtained from more than 108 current smokers, including more than 480 evaluable biopsy sites. The mean metaplasia index in these current smokers was 30.4%. For the total population studied, the median chromosome index for the bronchial biopsies was 1.41 (range, 1.04–1.61)

and the median chromosome polysomy index was 2.0% (range 0–8.7%). This can be compared to a mean chromosome index between 1.2–1.4 for lymphocytes and very rare chromosome polysomy. Interestingly, the intrasubject variability in chromosome instability was relatively low in most subjects and was less than the intersubject variability. These results suggested that chronic smokers harbor detectable chromosome instability throughout the accessible bronchial tree (supporting the field carcinogenesis notion) and that information from one biopsy site might yield representative information for the rest of the lung field.

Since most of the current smokers exhibited bronchial metaplasia in at least one of the biopsied sites, this allowed us to examine the relationship between chromosome instability and histologic changes, both on a site-by-site basis and on a per case basis. On a site-by-site basis, the chromosome indices of lesions showing squamous metaplasia were similar to those not showing metaplasia (i.e., median 1.43 vs. 1.43), and the degree of chromosome polysomy in metaplastic lesions were only slightly higher than in non-metaplastic sites (medians: 2.2% vs. 1.8%, respectively). Thus, the presence or absence of squamous metaplasia at a biopsy site does not necessarily correlate with the degree of underlying genomic instability. On the other hand, those subjects with metaplasia indices of at least 15% also showed higher levels of chromosome polysomy than did subjects with metaplasia index below 15% (medians: 2.4% vs. 1.8%,  $p = 0.005$ ). Thus, these chromosome instability assessments in current smokers appeared to reflect a more global process in the lung field.

Tobacco exposure has been shown to significantly increase the risk of developing lung cancer, and the degree of risk is related to the extent of tobacco exposure. We were interested in determining the relationship between individuals' smoking history parameters and the levels of chromosome change found in their lungs following years of tobacco exposure. While there was significant intersubject variation for similar tobacco exposure histories, overall there was a significant correlation between the degree of chromosome polysomy and the intensity of ongoing tobacco exposure (packs/day,  $p = 0.02$  on a per site basis) and with the extent of tobacco exposure (pack-years,  $p = 0.003$ ). Thus the amount of chromosome polysomy reflects the intensity and extent of tobacco exposure. At the same time, individuals with similar smoking histories showed widely divergent amounts of chromosome polysomy, possibly reflecting differences in intrinsic sensitivity between subjects. There was also strong correlation between the chromosome index and the duration of the smoking history (smoking years) and total accumulated exposure (pack-years,  $p = 0.0001$ ). These results suggest that tobacco exposure is associated with the initiation and accumulation of chromosome instability in the exposed lung; however individuals are differentially sensitive to carcinogenic insult. The working hypothesis is that those individuals who accumulate the highest degree of chromosome changes will be at the highest lung cancer risk.

Many of the bronchial biopsies from chronic smokers examined by *in situ* hybridization showed a rise in the chromosome index above that expected for a diploid cell population, especially in subjects with an extensive smoking history. The rise in chromosome index was also accompanied by an increase in the fraction of cells exhibiting at least 3 chromosome copies per cell. To determine if a rise in the tissue chromosome index was due to clonal expansion of populations with chromosome trisomy, the chromosome copy number and relative coordinates of each cell scored in

the bronchial epithelium was recorded and a spatial genetic map was created.<sup>41</sup> We then developed algorithms for calculating localized chromosome indices within the tissue. Since trisomic clones would have, on average, three chromosomes instead of two, those cells involved in neighborhoods with chromosome indices three-halves that of diploid populations could be marked as being part of a trisomic clone. Similarly, groups of cells with chromosome indices half that of diploid populations could be marked as being part of a monosomic clone. This allowed the generation of a second-order, two-dimensional genetic map representation of the bronchial epithelium showing the relative locations of cells involved in monosomic and trisomic clonal outgrowths. When adjacent tissue sections from the same bronchial biopsy were probed separately for different chromosomes, the detected clones appeared to occupy separate subregions of the epithelium. This result suggests that not only are the lungs of chronic smokers undergoing a process of genetic instability, they are experiencing the outgrowth of multiple clones throughout the exposed lung field, as postulated by the models shown in FIGURES 1 and 2. One advantage of this clonal approach is that the contribution of both monosomic and multisomic clones can be detected.

Since smoking cessation has been suggested to reduce the lung cancer risk, it was of interest to determine whether the levels of chromosome instability would decrease following smoking cessation. This question was possible to examine because our third sequential chemoprevention trial involved subjects who had discontinued smoking. So far, more than 220 subjects (more than 650 biopsies) who have quit smoking (mean 9.9 quit-years) have been evaluated for chromosome instability in their lungs. Despite the fact that the mean metaplasia index in this group is 5.8% (considerably less than that in current smokers), chromosome instability is still observed in the majority of subjects.<sup>42</sup> While the mean chromosome polysomy level is reduced to 1.0%, some individuals continue to show polysomy levels above 5%. Interestingly, while the overall chromosome polysomy levels were reduced in these individuals who stopped smoking, the mean chromosome index remained at about 1.4 with some individuals exhibiting chromosome indices as high as 1.8. Initial chromosome mapping studies suggest that while random chromosome instability seems to decrease following smoking cessation, the clonal outgrowths may remain for many years in the lung. The working hypothesis is that those individuals who show the greatest degree of remaining chromosome instability are at the highest lung cancer risk despite smoking cessation. Long-term follow-up on these subjects will be necessary to test this hypothesis.

## SUMMARY AND CONCLUSIONS

Aerodigestive tract tumorigenesis appears to be a multistep process taking place throughout the tissue fields of exposure. When viewed in the context of chromosome changes, carcinogen exposure appears to be associated with the random acquisition of chromosome polysomy throughout the exposed field, the degree of which is related to the degree and extent of carcinogen exposure as well as to the intrinsic susceptibility of the exposed individual. Continued exposure leads to continued acquisition of new changes and, in association with chronic wound-healing processes, to the

accumulation of clonal outgrowths throughout the target tissue. Although the ultimate malignancy may occur in only one or few tissue sites, manifestations of the instability process that drives tumorigenesis is globally present in the tissue. Thus random biopsies may provide useful risk information for the exposed field as a whole. Even when carcinogen exposure is reduced or chemopreventive strategies are initiated and histologic manifestations of the tumorigenesis process subside, the genetic scars of prior exposure remain in the form of clonal outgrowths and may explain continued lung cancer risk in ex-smokers. Future chemoprevention strategies need to focus on reducing the degree of chromosome instability and on trying to eliminate residual abnormal clonal outgrowths in the aerodigestive tract. In this setting, the measurement of chromosome instability in the target tissue will be useful in assessing cancer risk as well as response to intervention.

#### ACKNOWLEDGMENTS

The studies reviewed here represent one component of the collaborative efforts of the Aerodigestive Tract Chemoprevention team at The University of Texas M.D. Anderson Cancer Center, Houston, Texas. The studies were supported in part by National Institutes of Health-National Cancer Institute Grants CA 52051, CA 68437, CA 79437, CA 16672, CA 68089, CN 25433, CA 86390, CA 70907, NIH DE 13157, and the State of Texas Tobacco Research Fund.

#### REFERENCES

1. LANDIS, S.H., T. MURRAY, S. BOLDEN & P.A. WINGO. 1998. Cancer statistics, 1998. *CA Cancer J. Clin.* 48: 6-29.
2. JOHNSON, B.E. 1998. Second lung cancers in patients after treatment for an initial lung cancer. *J. Natl. Cancer Inst.* 90: 1335-1345.
3. LIPPMAN, S.M. & W.K. HONG. 1989. Second malignant tumors in head and neck squamous cell carcinoma. The overshadowing threat for patients with early stage of disease. *Int. J. Radiat. Oncol. Biol. Phys.* 17: 691-694.
4. SILVERMAN, S.J., JR., M. GORSKY & F. LOZADA. 1984. Oral leukoplakia and malignant transformation: a follow-up study of 257 patients. *Cancer* 53: 563-568.
5. LIPPMAN, S.M., J.S. LEE, R. LOTAN, *et al.* 1990. Biomarkers as intermediate endpoints in chemoprevention trials. *J. Natl. Cancer Inst.* 82: 555-560.
6. HEINONEN, O.P., D. ALBANES & THE ALPHA-TOCOPHEROL, BETA CAROTENE CANCER PREVENTION STUDY GROUP. 1994. The effect of vitamin E and beta carotene on the incidence of lung cancer and other cancers in male smokers. *N. Engl. J. Med.* 330: 1029-1035.
7. PETO, R., S. DARBY, H. DEO, *et al.* 2000. Smoking, smoking cessation, and lung cancer in the UK since 1950: combination of national statistics with two case-control studies. *Brit. Med. J.* 321: 323-329.
8. PERERA, F.P. 1996 Molecular epidemiology: insights into cancer susceptibility, risk assessment, and prevention. *J. Natl. Cancer Inst.* 88: 496-509.
9. LEE, J.S., S.M. LIPPMAN, S.E. BENNER, *et al.* 1994. Randomized placebo-controlled trial of isotretinoin in chemoprevention of bronchial squamous metaplasia. *J. Clin. Oncol.* 12: 937-941.

10. U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES. 1990. The health benefits of smoking cessation: a report of the Surgeon General. U.S. Department of Health and Human Services, Public Health Service, Centers for Disease Control, Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health. DHHS Pub. No. (CDC) 90-8416.
11. TONG, L., M.R. SPITZ, J.J. FAEGER, *et al.* 1996. Lung cancer in former smokers. *Cancer* 78: 1004-1010.
12. SLAUGHTER, D.P., H.W. SOUTHWICK & W. SMEJKAL. 1953. Field cancerization in oral stratified squamous epithelium: clinical implications of multicentric origin. *Cancer* 6: 963-968.
13. FARBER, E. 1984. The multistep nature of cancer development. *Cancer Res.* 44: 4217-4223.
14. CHUNG, K.Y., T. MUKHOPADHYAY, J. KIM, *et al.* 1993. Discordant p53 gene mutations in primary head and neck cancers and corresponding second primary cancers of the upper aerodigestive tract. *Cancer Res.* 53: 1676-1683.
15. SCHOLLES, A.G.M., J.A. WOOLGAR, M.A. BOYLE, *et al.* 1998. Synchronous oral carcinomas: independent or common clonal origin? *Cancer Res.* 58: 2003-2006.
16. GLUCKMAN, J.O., J.D. CRISSMAN & J.O. DONEGAN. 1980. Multicentric squamous cell carcinoma of the upper aerodigestive tract. *Head Neck Surg.* 3: 90-96.
17. AUERBACH, O., A.P. STOUT, E.C. HAMMOND, *et al.* 1961. Changes in bronchial epithelium in relation to cigarette smoking and in relation to lung cancer. *N. Engl. J. Med.* 265: 253-267.
18. SACCOMANNO, G., V.E. ARCHER, O. AUERBACH, *et al.* 1974. Development of carcinoma of the lung as reflected in exfoliated cells. *Cancer* 33: 256-270.
19. IZZO, J.G. & W.N. HITTELMAN. 1999. Characterization of multistep tumorigenesis by in situ hybridization. In *Introduction to Fluorescence In Situ Hybridization: Principles and Clinical Applications*. M. Andreeff & D. Pinkel, Eds.: 173-208. John Wiley & Sons, Inc. New York.
20. HITTELMAN, W.N. 1999. Molecular cytogenetic evidence for multistep tumorigenesis: implications for risk assessment and early detection. In *Molecular Pathology of Cancer*. S. Srivastava, D.E. Hensen & A. Gazdar, Eds.: 385-404. IOS Press. Amsterdam, The Netherlands.
21. SUNDARESAN, V., P. GANLY, R. HASLETON, *et al.* 1992. p53 and chromosome 3 abnormalities, characteristic of malignant lung tumours, are detectable in preinvasive lesions of the bronchus. *Oncogene* 7: 1989-1997.
22. KISHIMOTO, Y., K. SUGIO, J.Y. HUNG, *et al.* 1995. Allele-specific loss in chromosome 9p loci in preneoplastic lesions accompanying non-small-cell lung cancers. *J. Natl. Cancer Inst.* 87: 1224-1229.
23. CALIFANO, J., P. VAN DER RIET, W. WESTRA, *et al.* 1996. Genetic progression model for head and neck cancer: implications for field cancerization. *Cancer Res.* 56: 2488-2492.
24. PARK I.W., I.I. WISTUBA, A. MAITRA, *et al.* 1999. Multiple clonal abnormalities in the bronchial epithelium of patients with lung cancer. *J. Natl. Cancer Inst.* 91: 1863-1868.
25. SHIN, D.M., N. VORAVUD, J.Y. RO, *et al.* 1994. Sequential increases in proliferating cell nuclear antigen expression in head and neck tumorigenesis: a potential biomarker. *J. Natl. Cancer Inst.* 85: 971-978.
26. SHIN, D.M., J.Y. RO, W.K. HONG, *et al.* 1994. Dysregulation of epidermal growth factor receptor expression in premalignant lesions during head and neck tumorigenesis. *Cancer Res.* 54: 3153-3159.
27. SHIN, D.M., J. KIM, J.Y. RO, *et al.* 1994. Activation of p53 gene expression in premalignant lesions during head and neck tumorigenesis. *Cancer Res.* 54: 321-326.
28. IZZO, J.G., V.A. PAPADIMITRAKOPOULOU, X.Q. LI, *et al.* 1998. Dysregulated cyclin D1 expression early in head and neck tumorigenesis: in vivo evidence for an association with subsequent gene amplification. *Oncogene* 17: 2313-2322.
29. MAO, L., J.S. LEE, J.M. KURIE, *et al.* 1997. Clonal genetic alterations in the lungs of current and former smokers. *J. Natl. Cancer Inst.* 89: 857-862.

30. WISTUBA, I.I., S. LAM, C. BEHRENS, *et al.* 1997. Molecular damage in the bronchial epithelium of current and former smokers. *J. Natl. Cancer Inst.* 89: 1366-1373.
31. MAO, L., J.S. LEE, Y.H. FAN, *et al.* 1996. Frequent microsatellite alterations at chromosomes 9p21 and 3p14 in oral premalignant lesions and their value in cancer risk assessment. *Nature Med.* 2: 682-685.
32. PODDIGHE, P.J., F.C. RAMAEKERS & A.H. HOPMAN. 1992. Interphase cytogenetics of tumours. *J. Pathol.* 166: 215-224.
33. KIM, S.Y., J.S. LEE, J.Y. RO, *et al.* 1993. Interphase cytogenetics in paraffin sections of lung tumors by non-isotopic in situ hybridization. Mapping genotype/phenotype heterogeneity. *Am. J. Pathol.* 142: 307-317.
34. VORAVUD, N., D.M. SHIN, J.Y. RO, *et al.* 1993. Increased polysomies of chromosomes 7 and 17 during head and neck multistage tumorigenesis. *Cancer Res.* 53: 2874-2883.
35. HITTELMAN, W.N. 1999. Genetic instability assessments in the lung cancerization field. *In Lung Tumors: Fundamental Biology and Clinical Management.* C. Brambilla & E. Brambilla, Eds.: 255-267. Marcel Dekker. New York.
36. LEE, J.S., S.Y. KIM, W.K. HONG, *et al.* 1993. Detection of chromosomal polysomy in oral leukoplakia, a premalignant lesion. *J. Natl. Cancer Inst.* 85: 1951-1954.
37. LEE, J.J., W.K. HONG, W.N., HITTELMAN, *et al.* 2000. Predicting cancer development in oral leukoplakia: ten years of translational research. *Clin. Cancer Res.* 6: 1702-1710.
38. HITTELMAN W.N., R. YU, J. KURIE, *et al.* 1997. Evidence for genomic instability and clonal outgrowth in the bronchial epithelium of smokers [abstract]. *Proc. Am. Assoc. Cancer Res.* 38: 3097.
39. KURIE, J.M., J.S. LEE, F.R. KHURI, *et al.* N-(4-hydroxyphenyl)retinamide in the chemoprevention of squamous metaplasia and dysplasia of the bronchial epithelium. 2000. *Clin. Cancer Res.* 6: 2973-2979.
40. HITTELMAN, W.N., J.S. LEE, R.C. MORICE, *et al.* 1999. Lack of biomarker modulation in bronchial biopsies of chronic smokers following treatment with N-(4-hydroxyphenyl)retinamide (4-HPR). *Proc. Am. Assoc. Cancer Res.* 40: 2837.
41. HITTELMAN, W.N., J.S. LEE, N. CHEONG, *et al.* 1991. The chromosome view of "field cancerization" and multistep carcinogenesis. Implications for chemopreventive approaches. *In Chemoimmunoprevention of Cancer.* V. Pastorino & W.K. Hong, Eds.: 41-47. Georg Thieme Verlag. Stuttgart, Germany.
42. HITTELMAN, W.N., J.J. LEE, J.S. LEE, *et al.* 1998. Persistent genetic instability despite decreased proliferation in human lung tissue following smoking cessation. *Proc. AACR* 39: 336.